Department of Statistics

P-variation calculation and statistical inference for the discretely observed differential equation driven by fractional Brownian motion

A thesis submitted to the University of Warwick
for the degree of Doctor of Philosophy

Yang Zhao

Supervisor:
Dr. Anastasia Papavasiliou

July.2018

Contents

Acknowledgement

I would like to thank Dr.Papavasiliou for her help and advice for my research, and K.Taylor and T.Papamarkou for their academic help.

In addition, I would like to express my thanks to my dear friends Dialid, Xinya Guo, Yi Lu, Jessie Liu, Xiaofan Yang and Xin Tong who have been around to support me. At last, my parents, Abigale, Xiaoyuan, Yuxuan, Yanjun Jia, Jinru Han and Peixu Huang who though are far away have always been my source of strength.

Declaration

I declare that, to the best of my knowledge, all material contained in this thesis is my own original work,unless otherwise stated,cited or commonly known. No part of this thesis has been submitted for a degree to any other university.

Notation

$\|X\|_{p,J}$	The p-variation of path X on time interval J
$d_p(\mathbf{X}, \mathbf{Y})$	p-variation metric between two paths $\mathbf{X}$ and $\mathbf{Y}$
$E(X,p)_{[0,T]}$	The optimal partition for the p-variation of path X on time interval $[0,T]$
A^*	Transpose of matrix A
$diag(A)$	The diagonal elements of matrix A
$trace(A)$	The trace of matrix A
$\lambda_{max}(A)$	The maximum element of the eigenvalues of matrix A
$\|A\|_q$	The q-norm of matrix A
$A_\gamma \sim O(B_\gamma)$	$\lim_{\gamma\to\infty} \frac{A_\gamma}{B_\gamma} = C$ or $\lim_{\gamma\to 0} \frac{A_\gamma}{B_\gamma} = C$ with C a finite number
$A_\gamma \sim \mathrm{o}(B_\gamma)$	B_γ is an upper bound of A_γ when $\gamma \to \infty$ or $\gamma \to 0$
$\xrightarrow{P}$	Converge in probability P

Abstract

The dissertation is centred at the topic of parameter estimation for fractional stochastic differential equations. Three projects are included, and they all contribute to the problem of parameter estimation motivated by rough path theory. The first project derives an algorithm for the calculation of the p-variation of a piecewise linear path. One thing to be noticed is that the definition of p-variation is different from the definition in Itô's sense where the mesh of the partition goes to zero. The algorithm first transforms the path into a weighted graph and then exploits the optimal substructure.

The second one studies the method of the construction of approximate maximum likelihood estimators (MLEs) for discretely observed case by applying it to fractional Ornstein-Uhlenbeck (O-U) process. Parameter estimation for fractional processes is in the early development, especially for discretely observed cases. Most methods follow a strategy which approximates the exact likelihood function, and the method we study constructs the MLEs by building the exact likelihood function from an approximate model. The method is constructed under the frame work of rough path theory which makes it has a wide application even for the very rough paths.

The final project relates to the second one. We study an inverse algorithm in the final project which allows us to calculate the driving force $\{X_t\}$ given observations $\{Y_t\}$. The difficulty of the inverse problem we considered lies in the mismatch between data and model. We need to calculate the piecewise linear driving force of the approximate model, given the observations generated from the original continuous partial differential equation. The algorithm is motivated by the property of the signature of a path, that is, the signature is invariant by adding tree like paths. Thanks to the algorithm, the method in project two can be implemented to stochastic differential equations without analytical solutions, and we conduct numerical experiments on several such differential equations.

Chapter 1

Introduction

The dissertation is centred at the topic of parameter estimation for rough stochastic differential equations, and three projects are included. The first project derives an algorithm for the calculation of the p-variation of a path. The p-variation arising from Young's integral, is important in integral equation, operator analysis [23] and also in statistical inference. For instance, it can be used to estimate the volatility when the exact variation of a process is unknown. Despite many theoretical analysis involving p-variation, there are not many researches on the calculation of p-variation. Only one algorithm for the calculation is found. The algorithm uses segments and combine strategy which is different from our algorithm. We derive the algorithm by converting the path to a weighted graph and then exploit the property of optimal substructure.

The second one studies the method of the construction of approximate maximum likelihood estimators (MLEs) for discretely observed case by applying it to fractional Ornstein-Uhlenbeck (O-U) process. Brownian motion has been a popular random process to model random phenomenon, for instance in biology [66], and in [48], and thereby many studies contribute to the statistical inference of Gaussian process, especially when only partial observations can be obtained, since in most of the cases, we only have the observations at discrete time point in reality. In addition, the discretely observed case is more challenging because of the missing information which introduces error in our estimation. Many strategies for dealing with the error have been proposed. For the likelihood method, Ait-Sahalia [1] proposed a way to approximate the likelihood in a closed form. A computational approximation method using Monte-Carlo to simulate the approximate likelihood where the integrals in likelihood function are approximated by Riemann-Stiljes sum

has also been developed [54]. The method can be strongly biased if the time between observations is large [54]. Hence, the Martingale estimation method [42] is proposed in order to fix the bias problem by introducing a compensating factor. Apart from deriving the likelihood function, moment matching [25], Bayesian method [57] and indirect method are several among other strategies to solve the parameter estimation problem. A survey on methods tackling this issues can be found in [61].

Parameter estimation for fractional processes is in early development. For the continuously observed case, the methods are motivated by the fundamental martingale. Fractional process can be transformed to a fundamental martingale by applying some integral transformation to the process, and then usual martingale approach can be applied for the parameter estimation, see [35], and [37] for example. By the same virtue, Tudor et al [64]proposed another method for a general class of fractional differential equations, which is by far the most general method for the continuously observed cases for fractional processes. Partial discussion for discretely observed case is made in [64]. The above methods follow the same strategy which constructs the estimators by approximating the likelihood function of the continuous model, and the method we study constructing the MLEs by building the exact likelihood function for an approximate model. The approximate model is a piecewise smooth ordinary differential equation, in comparison to the integral transformation in fractional calculus, it is easy to understand and to apply.

For the above method to work, we need to calculate the driving force $\{X_t\}$ given observations $\{Y_t\}$, and we study such an algorithm in the final project. Inverse problem for differential equation can be loosely characterised as follows, given observations $\{Y_t\}$ satisfying$Y_t = F(\theta, u_t)$, we find the function $\{u_t\}$. The difficulty lies in the fact that the inverse operator F^{-1} is not necessarily well defined, and thus many studies focus on the analysis of the regularity of the operator F^{-1} [31]. The inverse problem we considered is different from the above for the following reasons, we assume sufficient regularity of the inverse operator, and the difficulty lies in the mismatch between data and model. We need to calculate the piecewise linear driving force of the approximate model, given the observations generated from the original continuous partial differential equation. The algorithm is motivated by the property of the signature of a path, that is, the signature is invariant by adding tree like paths. Thanks to the algorithm, the method in project two can be implemented to stochastic differential equations without analytical solutions, and we conduct numerical experiments on several such differential equations.

The thesis is organized as follows. In Chapter 2, we provide some preliminaries on rough path theory which relates to our projects. Chapter 3 is about the algorithm of the calculation of the p-variation, Chapter 4 is devoted to the parameter estimation problem for the discretely observed fractional processes, and finally, the inverse algorithm is discussed in Chapter 5. In addition, a list of theorems which will be used are presented in the Appendix for the reference of the readers. What's more, some calculations and proofs for the second project are also included in the Appendix.

Chapter 2

Preliminaries on rough path theory

In this chapter, some important and relating concepts in rough path theory are introduced, with an focus on p-variation, signature and the rough differential equation. Most of the material can be found in [41], and thus we will not make explicit citation unless it is from other source.

2.1 P-variation

The p-variation is a basic yet important concept in rough path theory. In this section, we will introduce some properties of the p-variation. The p-variation is defined as follows,

Definition 1. *[41] Let $\{X_t\} : J \to E$ be a continuous path where $(E, |.|)$ is a Banach space, and $p \geq 1$ be a real number. Then the p-variation of $\{X_t\}$ on interval J is*

$$||X||_{p,J} = \left[\sup_{\mathcal{D} \subset J} \sum_j |X_{t_{j+1}} - X_{t_j}|^p \right]^{\frac{1}{p}}.$$

It is often referred to as the true p-variation, in order to distinguish it from the p-variation in Itô's definition. The difference is that, the quadratic variation is calculated by letting the mesh goes to zero, whereas p-variation it is calculated by taking the supreme of the p-variation over all the possible finite partitions, and the mesh does not go to zero. It is obvious that $||X||_{p,J} \geq |X_t - X_s|$, and therefore a path is of zero p-variation if and only if it is a constant path. Generally, an α-Hölder

continuous path has finite true $\frac{1}{\alpha}$-variation. Therefore Brownian motion has finite non-zero p-variation for any $p > 2$, and its 2-variation is unbounded. However, for the partition whose mesh goes to zero, the 2-variation of Brownian motion is T for any interval [0,T] almost surely. For any $p > 2$, such partition will yield zero p-variation of Brownian motion, but the true p-variation will be a positive finite number. Next, two basic but important properties of the true p-variation will be given which were proposed by T.Lyons [41] (Chapter 1.2, p4, Lemma 1.6).

Property 1. *Let $\varphi : J \to J$ be a non-decreasing surjection. Then $||X||_{p,J} = ||X \circ \varphi||_{p,J} \quad \forall p \geq 1$.*

Property 2. *The function $p \to ||X||_{p,J}$ from $[0, \infty)$ to $[0, \infty]$ is non-increasing.*

2.2 Signature

Signature is an important concept in rough path theory. Signature, defined as a series of iterated integral over a path, provides us with an efficient way of summarizing the features of quickly oscillating paths. Traditionally, paths are taken as an evolving system over time horizon, and the way to deal with them is to sample them at some discrete interval. However, the method can be inefficient when the paths are quickly oscillating, since it relies on sampling them at a very fine grid so that we would not omit the picks over small intervals. What's more, we can see that redundant information might occur for sampling method, due to the repeated behaviour of paths over certain time interval. We can ask the question that whether it is necessary to include all the picks at a very fine grid, and how much new information will we gain from the finer and finer sampling. On the contrary, the signature of a path provides us with a tool to compress the data without losing the key features. The paths can be broken down into several segments and we can represent the features of the path by its signature under each segment. It is proved to be a more efficient way of summarising an evolving system than sampling the system at fine grid for Brownian motion case[19]. In addition, it has many successful applications, such as to financial market data [40], Chinese handwriting recognition [28], and medical and neuroimaging pattern recognition [13].

Let $X : J \to E$ be a n-dimensional continuous path, with J a compact interval, and E a Banach space. The i-th iterated integral of X has the following expression

$$X^i_{s,t} = \int \int \cdots \int_{s \leq t_1 < t_2 < \ldots < t_i \leq t} dx_{t_1} \otimes dx_{t_2} \otimes \ldots \otimes dx_{t_i}$$

Each i-th iterated integral lives in a tensor space $E^{\otimes i}$ with dimension N^i where N is the dimension of E.
It should be noticed that the definition of the iterated integral has some ambiguity as to the definition of the integral, and we will not elaborate on this issue because it is not the focus of this report. The integral can be defined, for example, in Ito's sense or Stronovich's sense as long as they satisfy the Chen's Identity

$$X_{s,t}^k = \sum_{i=0}^{k} X_{s,u}^i \otimes X_{u,t}^{k-i}$$

The relevant discussion can be found in Chapter 3 and 4 in T.Lyon's St.Flour notes[41].

Definition 2. *We denote $S(X)$ to be the signature of the d-dimensional continuous path X, and thus*

$$S(X) = (1, X^1, X^2, ..., X^n, ...)$$

$S(X)$ is an object which contains all levels of the iterated integral of X, and $S(X)$ lives in tensor space $T(E) = \bigoplus_{i=0}^{\infty} E^{\otimes i}$. An important operation on the signature is the concatenation,

$$S(X)_{s,t} = S(X)_{s,u} \otimes S(X)_{u,t},$$

and we define the inverse $S(\overleftarrow{X})_{u,s}$ of a signature $S(X)_{u,s}$ as follows

$$\mathbf{1} = S(X)_{s,u} \otimes S(\overleftarrow{X})_{u,s}, \quad \mathbf{1} = S(\overleftarrow{X})_{u,s} \otimes S(X)_{s,u},$$

where $\mathbf{1} = \{1, 0, 0, ...\}$ the unit element, is the signature of a constant path. As we can see that the signature of paths forms a group over concatenation operation.
Another important property of signature is invariance over time parametrization. Consider a path $X : [0, T] \to E$, and another path $\tilde{X}_{\tau(t)}$ under the time parametrization $\tau : [0, T] \to [0, T]$. Then we have

$$S(X)_{[0,T]} = S(\tilde{X})_{[0,T]}.$$

As we can see the first few levels of iterated integral contains most of the information, and we can actually truncate the signature of a p-rough path up to $\lfloor p \rfloor$ level. We define $T^{\lfloor p \rfloor} = \pi_{\lfloor p \rfloor}(T(E))$ to be the truncated tensor space.

Let $C_{0,p}(\Delta_T, T^{\lfloor p \rfloor}(E))$ be a functional space which contains all the continuous functions from the simplex Δ_T to the truncated tensor algebra $T^{\lfloor p \rfloor}(E)$ with finite p-variation, and let $\mathbf{X}, \mathbf{Y} \in C_{0,p}(\Delta_T, T^{\lfloor p \rfloor}(E))$. Then the p-variation metric between two paths $\mathbf{X}$ and $\mathbf{Y}$ is defined as follows,

$$d_p(\mathbf{X}, \mathbf{Y}) = \max_{1 \leq i \leq \lfloor p \rfloor} \sup_{\mathcal{D}} \left(\sum_{\mathcal{D}} ||X^i_{t_{l-1},t_l} - Y^i_{t_{l-1},t_l}||^{\frac{p}{i}} \right)^{\frac{i}{p}}$$

Other important properties on the signature include shuffle properties, and log signature, an important transformation of the signature. Please refer to the St.Flours notes for more information, or papers[40], [28], and [13] which are about application of the properties of the signatures.

2.3 Rough Differential Equation

We give the definition of the p-rough path first in order to give the definition of the rough differential equation. Beforehand, we introduce the definition of multiplicative functional.

Definition 3. *(p42, Section 3.2.1, Definition 3.1) Let V be a Banach space, and $n \geq 1$ be an integer. Let Δ_T be a simplex on $[0, T]$, and $T^n(V)$ be an n^{th} level of truncation of the signature space of V. Let $\mathbf{X} : \Delta_T \to T^n(V)$ be a continuous map. For each $(s, t) \in \Delta_T$, denote the image of (s, t) by $X_{s,t}$, and write*

$$\mathbf{X}_{s,t} = (\mathbf{X}^0_{s,t}, \mathbf{X}^1_{s,t}, ..., \mathbf{X}^n_{s,t}) \in \mathbb{R} \oplus V \oplus V^{\otimes 2} \oplus \cdots \oplus V^{\otimes n}.$$

The function $\mathbf{X}$ is called a multiplicative functional of degree n in V if $\mathbf{X}^0_{s,t} = 1$ for all (s, t), $(t, u) \in \Delta_T$, and

$$\mathbf{X}_{s,t} \otimes \mathbf{X}_{t,u} = \mathbf{X}_{s,u}.$$

Next, we give the definition of p-rough path.

Definition 4. *(p52, Section 3.2.1, Definition 3.11) Let V be a Banach space. Given a real number p, $\mathbf{X}$ is a $\lfloor p \rfloor$-rough path if it is a multiplicative functional of degree $\lfloor p \rfloor$ in V and has finite $\lfloor p \rfloor$-variation.*

Please refer to Definition 1 for the definition of $\lfloor p \rfloor$-variation. As we can see, the first level of the rough path is the increment of the original path. A geometric p-rough path is a p-rough path that can be expressed as the limit of 1-rough paths in

the p-variation metric. For two p-rough paths, $\mathbf{X}$, $\mathbf{Y}$, and a Banach space $(V, \|.\|)$, the p-variation metric is defined as follows (p52, Section 3.2.1)

$$d_p(\mathbf{X}, \mathbf{Y})_{[0,T]} = \max_{i=1,..,\lfloor p \rfloor} \sup_{\mathcal{D}} (\sum_{j \in \mathcal{D}} \|X^i_{t_j,t_{j-1}} - Y^i_{t_j,t_{j-1}}\|^{\frac{p}{i}})^{\frac{i}{p}}, \tag{2.1}$$

where $\mathcal{D}$ is the time partition of $[0,T]$. The space of geometric p-rough path in V is denoted by $G\Omega_p(V)$.
We then introduce the universal limit theorem which guarantees that the solution of a differential equation will be a geometric rough path when the driving force is a geometric rough path.

Theorem 2.3.1 (Universal Limit Theorem)**.** *(p83, Section 5.3, Theorem 5.3) Let V and W be two Banach spaces. Let $p \geq 1$ and $\gamma > p$ be real numbers. Let $f : W \to L(V,W)$ a γ-Lipschitz continuous function. For all $X \in G\Omega_p(V)$ and $\eta \in W$, the equation*

$$dY_t = f(Y_t)dX_t, \quad Y_0 = \eta$$

admits a unique solution which depends continuously on X and η in p-variation metric. The rough path Y is a p-geometric rough path.

We give the definition of the γ Lipschitz function given in the book [27] by P.Friz and N.Victoir.

Definition 5. *(P216, Chpter 10.1, Definition 10.2) A map $Q : \mathbb{E} \to F$ between two normed space E, F is γ-Lipschitz if Q is $\lfloor \gamma \rfloor$ times continuously differentiable and there exists a M, such that the supreme norm of its k-th derivative, $k = 0, 1, ..., \lfloor \gamma \rfloor$ and the $\{\gamma\}$-Hölder norm of its $\lfloor \gamma \rfloor$ derivative is bounded by M, where $\{\gamma\} = \gamma - \lfloor \gamma \rfloor$.*

Chapter 3

Optimal partition for p-rough paths

In the first project, an algorithm for calculating the p-variation of piecewise linear paths is derived. Given an n-piecewise linear path, there are $1 + \sum_{k=1}^{n-2} C_{n-2}^{k}$ number of combinations, which means that the number of possible combinations grows exponentially with respect to n, and thus finding the optimal partition is computational intensive.

Despite the difficulty, it is worthwhile calculating the p-variation of a sample path since it can be applied to the parameter estimation problem as a more robust estimator than the quadratic variation. Because by definition, p-variation can never be zero, and as long as p is large enough, the p-variation is finite. For example, the empirical financial data [26], [34] suggests that the stock price does not follow a semi-martingale process if we consider the process on a fine time grid, and thus as we take the observations on a very fine grid, the quadratic variation does not converge which renders the quadratic variation useless or force us to only take observations at a rougher time grid. On the contrary, p-variation as an estimator can yield something meaningful since we can adjust the magnitude of p in order to for p-variation to be a finite number rather than having an infinite estimator as quadratic variation. For instance, for a path of finite 2.1 variation, its quadratic variation will go to infinity, but we can find a finite p-variation of the path as long as $p > 2.1$. By the property, when estimating the diffusion coefficient of the limiting homogenized multiscale model, the use of the p-variation can avoid the problem of subsampling[50] which exists for the estimator derived by the quadratic variation.

In addition, motivated by [50], T. Manikas and A.Papavasiliou [63]proposed an estimator which is the quadratic variation [53]corresponding to the local extrema of the process. This estimator is easier to compute than the estimator in [50], and the L_2 error is smaller than the estimator using the quadratic variation [63].
The chapter is organized as follows. First, we provide some literature review on the topic of p-variation calculation, and then, we provide necessary background for the algorithm, which is optimal substructure. Since we will conduct numerical experiments on the diffusion coefficient estimation for fractional Brownian motions as an example of the application of the p-variation, we also provide the background on fractional Brownian motion. Then we explain the algorithm for one-dimensional and multi-dimensional cases and then we conduct some numerical experiments.

3.1 Literature review

In paper[22], Driver discusses the problem of calculating the p-variation of a piecewise linear path $\{X_t\}$ on $\{t_1, t_2, ..., t_n\}$, and proves the result that the optimal partition is a subset of the point set of the extrema. Extrema point X_{t_i} is a point such that $X_{t_i} > max(X_{t_{i-1}}, X_{t_{i+1}})$, or $X_{t_i} < min(X_{t_{i-1}}, X_{t_{i+1}})$. The author considers one kind of path called jog-free, and proves that if the path is jog free, then the optimal partition includes all the extrema points. The jog free path is defined as follows,

Definition 6. *Let $\{X_t, t \in [0,T]\}$ be a path which only contains extrema points, and if the extrema points satisfy the property where all local maximum points X_t are greater than X_s for $s > t$, and all local minimum points X_t are smaller than X_s for $s > t$, then the path is called jog free.*

Though no complete algorithm is given in [22], the results obtained by Driver provide the starting point for further research. An algorithm for calculating p-variation of a one-dimensional sample path is given by V.Butkus [11]. The main idea is to first find optimal partitions for some sub intervals and then combine the sub intervals together. The advantage about the algorithm is that by separating the path into several segments, we can use parallel computing to simultaneously find the optimal partition for each segment, which accelerates the speed. However, the algorithm dose not simplify the procedure for finding the optimal partition in a sense that the algorithm searches every possibility of the partition when calculating the optimal partition for each segment. Moreover, the combining procedure is also

calculation intensive, since we have to consider all the points from both intervals in order to find the joint points defined as the adjacent points which originally come from two different segments. The procedure is repeated for every segment of the path till we have the optimal partition for the whole path. As we can see that the algorithm does many redundant work by searching through all the possible combinations.
The algorithm proposed by V.Butkus [11] solves the problem using a dynamic algorithm, in which the optimal problem is decomposed into several sub problems, and then combining the sub problems to achieve the optimization. While our method tackles the problem from the greedy algorithm point of view, that is we search for the optimal partition along the time points in sequence and in the end, we have the optimal partition for the path on the whole interval. Our algorithm calculates the optimal partition for the time being (local optimization) which is the characterization of greedy algorithm.

3.2 Preliminaries

In this section, we introduce some background knowledges on optimal substructure and fractional Brownian motion.

3.2.1 Optimal substructure

In order to find the optimal partition, we first transform the path into a graph, and thus the problem becomes an optimization problem for the graph. Thus, we provide some background on graph theory. By [21](Chapter 1.1, p2), a graph is a pair $G = (V, E)$ of sets such that $E \subseteq [V]^2$, and a directed graph is defined as follows,

Definition 7. *A directed graph is a pair (V, E) of disjoint sets (of vertices and edges) together with two maps $init : E \to V$ and $ter : E \to V$ assigning to every edge e an initial vertex $init(e)$ and a terminal vertex $ter(e)$. The edge e is said to be directed from $init(e)$ to $ter(e)$.*

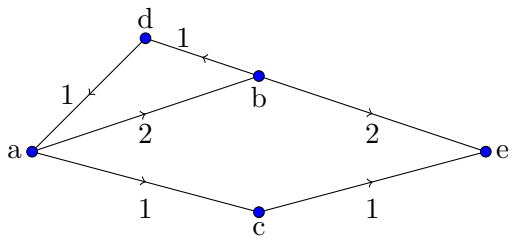
A path of a graph G, is a pair from (V, E), such that $V = \{v_0, v_1, ..., v_m\}$, and $E = \{v_0v_1, v_1v_2, ..., v_{m-1}v_m\}$. If $v_m = v_0$, then the path is a cycle. A directed graph is said to be a directed acyclic graph (DAG) if it has no cycle. Every edge can have a weight, and a weight of edge v_iv_j is denoted by $w(v_iv_j)$. For a weighted directed graph (a directed graph assigning each edge a weight), an important problem is the

searching for the longest path. Generally, the search for the longest path in a graph is an NP-hard problem[59], but since DAG has the optimal substructure property, the longest path problem for DAG can be solved in time $O(|V| + |E|)$ [16] where $|V|$ is the number of vertices and $|E|$ is the number of edges of the graph.
Optimal substructure refers to the property that the optimal property (longest, or shortest) of a path is preserved by any sub-path[16]. For example, if $(v_1, v_2, ..., v_n)$ is the shortest path from v_1 to v_n in a graph G, then any sub-path $(v_i, v_{i+1}, ..., v_{i+m})$ is the shortest path from v_i to v_{i+m} in G. We prove that the longest path problem for DAG has optimal substructure in the following,

Lemma 3.2.1. *For a weighted directed acyclic graph $G = (V, E)$, the longest path problem for G has optimal substructure.*

Proof. Suppose $(v_1, v_2, ..., v_n)$ is the longest path in a weighted DAG, and if there exists a sub-path of the longest path, say $(v_i, v_{i+1}, ..., v_{i+m})$ such that the longest path from v_i to v_{i+m} is not $(v_i, v_{i+1}, ..., v_{i+m})$, then by swapping the segment, we can have a longer path from v_1 to v_n. This contradicts the assumption that $(v_1, v_2, ..., v_n)$ is the longest path. □

The reason we can do the swap is that the graph has no cycle. Thus, any swap would yield a path. We provide a counter-example to illustrate the point.



From the above graph, the longest path from a to e is $\{a, b, e\}$, but the longest path from b to e is not $\{b, e\}$. The longest path from b to e is $\{b, d, a, c, e\}$, and thus, we can see that graphs with cycle do not have the optimal substructure.

3.2.2 Fractional Brownian motion

We provide some background on fractional Brownian motion, and the material can be found in [46]. We consider $\{B_t^h\}$ on $(\Omega, \mathcal{F}, \mathcal{F}_{(t>0)}, P)$ with Hurst index h. The one-dimensional fractional Brownian motion is a continuous Gaussian process

with zero expectation. However, it does not necessarily have independent increment in contrast to standard Brownian motion. The covariance between B_t and B_s is

$$\mathbb{E}(B_t^h B_s^h) = \frac{1}{2}\{t^{2h} + s^{2h} - |t-s|^{2h}\}$$

where $h \in (0,1)$ is the Hurst index. Hence the covariance of the increment is

$$\mathbb{E}((B_{t+r}^h - B_t^h)(B_{s+r}^h - B_s^h)) = \frac{1}{2}\{|t-s+r|^{2h} + |t-s-r|^{2h} - 2|t-s|^{2h}\}.$$

The above expression indicates that the increment of fractional Brownian motion has stationary distribution. In addition, we can see from the above equation that fractional Brownian motion is Brownian motion when $h = 0.5$. When $h > 0.5$, the increments are positively correlated, and fractional Brownian motion has long time memory. On the contrary, when $h < 0.5$, the increments are negatively correlated, and therefore the process oscillates more quickly .
Consider a random vector of increments of fractional Brownian on a homogeneous time point set with $t_i - t_{i-1} = \delta$,

$$\Delta B^h = [\Delta B_{t_1}^h, \Delta B_{t_2}^h, \Delta B_{t_3}^h, ..., \Delta B_{t_n}^h]^*,$$

where $\Delta B_{t_{i+1}}^h = B_{t_{i+1}}^h - B_{t_i}^h$, and the covariance matrix of ΔB^h is as follows

$$\Sigma_n = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & \rho_0 \end{bmatrix}, \tag{3.1}$$

with

$$\begin{aligned} \rho_{|i-j|} &= \mathbb{E}(\Delta B_{t_i}^h \Delta B_{t_{n-j}}^h) \\ &= \frac{\delta^{2h}}{2}\{(|i-j|+1)^{2h} + (|i-j|-1)^{2h} - 2|i-j|^{2h}\}. \end{aligned}$$

As $|i-j| = m >> 1$, we have [46]

$$\rho_m = \frac{\delta^{2h}}{2} O(m^{2h-2}). \tag{3.2}$$

3.3 One-Dimensional case

We would like to explain the one-dimensional case in this section. We first introduce some notation, and then we provide a representation of the problem in terms of DAG.

3.3.1 Setting

Notation Let $\{X_t^\delta, t \in [0,T]\}$ be the piecewise linear interpolation of a real valued path $\{X_t\}$ with finite p-variation on homogeneous time point set

$$\mathcal{D}_0 = \{t_0, t_1, ..., t_n\} = \{0, \delta, 2\delta..., T\},$$

which we call the original partition. We give some definitions and notations which will be referred to frequently.

- Optimal Partition. The partition we use to calculate the p-variation of $\{X_t^\delta\}$ on $[0,T]$, denoted by $E(X^\delta, p)_{[0,T]}$.
- Monotonic Sequence. If the point set $X^\delta = \{X_{t_i}^\delta, X_{t_{i+1}}^\delta, ..., X_{t_k}^\delta\}$ is monotonically decreasing or increasing, then we call the point set X^δ a monotone sequence and we also say that the points are in the same direction.
- Corner Point. $X_{t_i}^\delta$ is a corner point, if $\{X_{t_{i-1}}^\delta, X_{t_i}^\delta, X_{t_{i+1}}^\delta\}$ is a monotonous sequence.
- Extremal Point. $X_{t_i}^\delta$ is an extremal point if $X_{t_i}^\delta > max(X_{t_{i-1}}^\delta, X_{t_{i+1}}^\delta)$ or $X_{t_i}^\delta < min(X_{t_{i-1}}^\delta, X_{t_{i+1}}^\delta)$ where $X_{t_{i-1}}^\delta$ and $X_{t_{i+1}}^\delta$ are the neighbour points of $X_{t_i}^\delta$ in a partition
- Extremal Partition. Extremal partition is the partition which contains only the extremal points and is denoted by $\mathcal{D}_m(X^\delta)_{[0,T]}$.
- Pre-p-variation. The pre-p-variation is defined as $V_p(\mathcal{A}, X^\delta) = (\sum_{t_i \in \mathcal{A}} |X_{t_{i+1}}^\delta - X_{t_i}^\delta|^p)^{1/p}$, where $\mathcal{A}$ is a finite partition of some time interval.

It is worth noticing that the extremal partition is not unique for a specific given partition, and a piecewise linear path can have more than one extremal partition. In our case, we consider the extremal partition with respect to the original partition

$\mathcal{D}_0$, which is denoted as $\mathcal{D}_m$. We illustrate the type of points by the following picture.

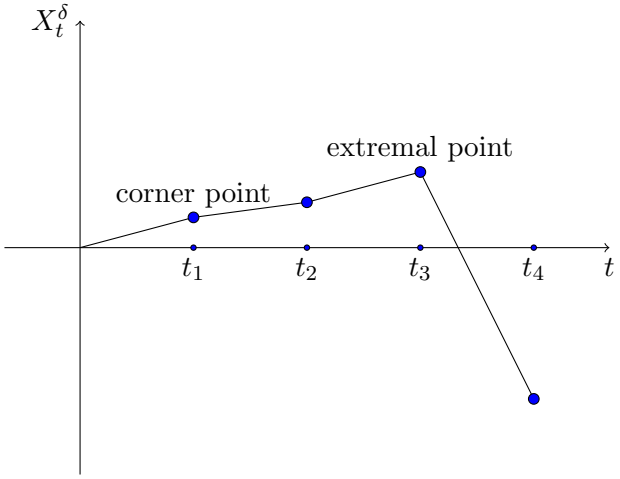
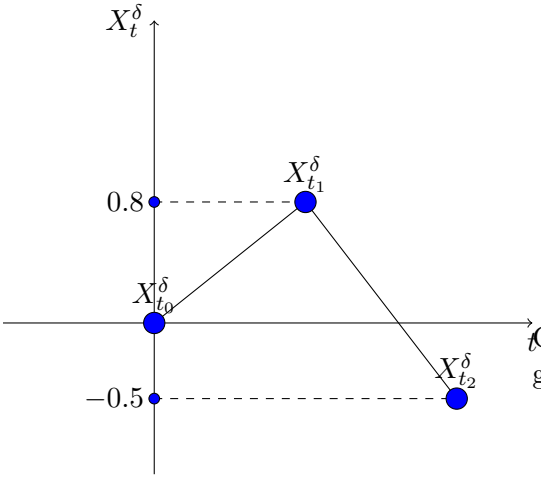
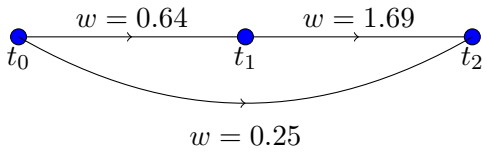


Figure 3.1: Category of points

Representation We would like to transform the piecewise linear path into a weighted directed graph, and thus the problem of finding the optimal partition can be solved by finding the longest path in the graph. Actually, the piecewise linear path can be transformed to the directed acyclic graph (DAG). If t_u and t_v are two vertices with $u < v$, then we draw a directed edge from t_u to t_v, and in this way we can turn the path into a DAG. Since time has only one direction, the graph has no cycle. The search for the optimal partition is equivalent to the search of the longest path of the DAG from t_0 to t_n where each edge is weighted in the following way $w(t_u t_v) = |X_{t_u}^\delta - X_{t_v}^\delta|^p$.



Piecewise linear path



Corresponding weighted directed acyclic graph

Since the longest path problem of a DAG has the optimal substructure by Lemma 3.2.1, the optimal partition has optimal substructure by the above transformation. The substructure of optimality guarantees the algorithm of searching for the optimal partition consumes linear time, and the complexity of the algorithm will be discussed in Section 2.3.3. By the substructure of optimality, the optimal partition of $\{X_t^\delta, t \in [0, t_i]\}$ is $E(X^\delta, p)_{[0,T]}$ truncated at time t_i, denoted by $E(X^\delta, p)_{[0,T]}|_{[t_0,t_i]}$. Also, the optimal partition of $\{X_t^\delta, t \in [t_i, T]\}$ is $E(X^\delta, p)_{[0,\tau_{i+1}]}|_{[t_i,T]}$. We make a theorem about the optimal substructure of the optimal partition.

Theorem 3.3.1. *$E(X^\delta, p)_{[0,T]}$ has optimal substructure property.*

Proof. Since we can transform the path to a DAG, and the searching for the optimal partition is equivelent to finding the longest path for the DAG. From Lemma 3.2.1, a DAG has the optimal substructure property, and so does the optimal partition for a given path. □

3.3.2 P-variation algorithm

Before presenting the algorithm, we explain the components of the algorithm.

Extremal point set The first theorem we want to prove is that the optimal partition of p-variation for $\{X_t^\delta\}$ is a subset of the extremal partition $\mathcal{D}_m$, the extremal partition with respect to the original partition $\mathcal{D}_0$. The result is also proven by Driver [22]. Beforehand, we prove a lemma.

Lemma 3.3.1. *Let $\{X_t^\delta\}$ be a 1-dimensional piecewise linear path on $[0,T]$. For a partition $\mathcal{A} \subseteq \mathcal{D}_0$, if there exists $H = \{t_i, t_{i+1}, ..., t_{i+m}\} \subseteq \mathcal{A}$ such that $\{X_t^\delta\}$ is a monotonic sequence on H. Then by excluding $\{t_{i+1}, ..., t_{i+m-1}\}$ from $\mathcal{A}$, the pre-p-variation of $\{X_t^\delta\}$ can be increased where $p \geq 1$. That is,*

$$V_p(\mathcal{A}\backslash\{t_{i+1}, t_{i+2}, ..., t_{i+m-1}\}, X^\delta) \geq V_p(\mathcal{A}, X^\delta)$$

Proof. Let $a_i = X_{t_{i+1}}^\delta - X_{t_i}^\delta$ and $c = X_{t_{i+m}}^\delta - X_{t_i}^\delta$. Since $\{X_t^\delta\}$ is monotonic on $\{t_i, t_{i+1}, ..., t_{i+m}\}$, we have $|c| > |a_j|$ for $j \in \{i, i+1, ..., i+m-1\}$, and

$$\sum_{j=i}^{m-1} |a_j| = |c|. \qquad (*)$$

Since $p \geq 1$, by multiplying both sides of equation (*) by $|c|^{p-1}$, we obtain

$$|c|^p = |c|^{p-1} \sum_{j=i}^{m-1} |a_j| \geq \sum_{j=i}^{m-1} |a_j|^p.$$

That is

$$|X_{t_{i+m}}^\delta - X_{t_i}^\delta|^p \geq \sum_{j=i}^{m-1} |X_{t_{j+1}}^\delta - X_{t_j}^\delta|^p.$$

By the definition of p-variation, the lemma is proved. □

We prove the following theorem by the above lemma, and the theorem is first proved by Diver [22].

Theorem 3.3.2. *Let $\{X_t^\delta\}$ be a 1-dimensional piecewise linear path. The optimal partition of the p-variation of $\{X_t^\delta\}$ is a subset of its extremal partition. That is $E(X^\delta, p)_{[0,T]} \subset \mathcal{D}_m$.*

Proof. Let $\mathcal{A} \subseteq \mathcal{D}_0$ be a partition, and let $t_i \in \mathcal{A}$ be corner point with respect to the original partition $\mathcal{D}_0$. Then there are two cases for $X_{t_i}^\delta$ with respect to $\mathcal{A}$. First, $X_{t_i}^\delta$ is a corner point in $\mathcal{A}$. By Lemma 3.3.1, we should exclude $X_{t_i}^\delta$. Secondly, $X_{t_i}^\delta$ becomes an extremal point in $\mathcal{A}$. Suppose $X_{t_i}^\delta > max(X_{t_{i-1}}^\delta, X_{t_{i+1}}^\delta)$ where $X_{t_{i-1}}^\delta$ and $X_{t_{i+1}}^\delta$ are the neighbour points of $X_{t_i}^\delta$ in $\mathcal{A}$. Because $X_{t_i}^\delta$ is not an extremal point in $\mathcal{D}_0$, we can find a point X_τ^δ such that $X_\tau^\delta > X_{t_i}^\delta$ with $t_{i-1} < \tau < t_{i+1}$. Then, we have

$$|X_\tau^\delta - X_{t_{i-1}}^\delta|^p + |X_\tau^\delta - X_{t_{i+1}}^\delta|^p > |X_{t_i}^\delta - X_{t_{i-1}}^\delta|^p + |X_{t_i}^\delta - X_{t_{i+1}}^\delta|^p$$

Therefore, $\mathcal{A}$ cannot be the optimal partition for the p-variation of $\{X_t^\delta, t \in [0,T]\}$. □

Second last point By optimal substructure, we prove that the optimal partition has a good iterative property. That is, $E(X^\delta,p)_{[0,\tau_{i+1}]} = E(X^\delta,p)_{[0,\tau^*]} \bigcup \tau_{i+1}$, where $\tau^* \in E(X^\delta,p)_{[0,\tau_i]}$, and τ^* is defined as the second last point. We would like to derive the corollary from Theorem 3.3.1.

Corollary 1. *The optimal partition for* $\{X_t^\delta, t \in [0,\tau_{i+1}]\}$ *is of the form* $E(X^\delta,p)_{[0,\tau_{i+1}]} = E(X^\delta,p)_{[0,\tau^*]} \bigcup \tau_{i+1}$, *where* $\tau^* \in E(X^\delta,p)_{[0,\tau_i]}$.

A lemma is presented first which will be useful for the proof of the above corollary .

Lemma 3.3.2. *Let* $x,\ \ y > 0$, *and* $x + \sum_i \alpha_i + y > 0$. *If* $(x + \sum_i \alpha_i + y)^p > x^p + \sum_i |\alpha_i|^p + y^p$, *then*

$$(a + x + \sum_i \alpha_i + y + b)^p > |a+x|^p + \sum_i |\alpha|^p + |b+y|^p,$$

where $a,\ \ b > 0$, *and* $p > 1$.

Proof. Define function $f(x,y) = (x + \sum_i \alpha_i + y)^p - x^p - \sum_i |\alpha_i|^p - y^p$, and we have

$$\frac{\partial f(x,y)}{\partial x} = p((x + \sum_i \alpha_i + y)^{p-1} - x^{p-1}),$$

$$\frac{\partial f(x,y)}{\partial y} = p((x + \sum_i \alpha_i + y)^{p-1} - y^{p-1}).$$

Since $(x+\sum_i \alpha_i+y)^p > x^p+\sum_i |\alpha_i|^p+y^p$, $\frac{\partial f(x,y)}{\partial x}$ and $\frac{\partial f(x,y)}{\partial y}$ are positive, and hence $f(x,y)$ is increasing with respect to each variable x and y. In addition, a and b are positive number, $f(a+x,b+y) > f(x,y) > 0$. □

We proceed to prove the corollary 1.

Proof. Suppose $E(X^\delta,p)_{[0,\tau_{i+1}]}$ includes points which do not belong to $E(X^\delta,p)_{[0,\tau_i]}$, and let $X^\delta_{t'_k}$ be the last of such point. Then, it is the second last point in $E(X^\delta,p)_{[0,\tau_{i+1}]}$ by the optimal substructure. Let $X^\delta_{\tau'_{k-1}}$ and $X^\delta_{\tau'_{k+1}}$ be the closest left and right points to $X^\delta_{t'_k}$ from $E(X^\delta,p)_{[0,\tau_i]}$. Then $\{X^\delta_{\tau'_{k-1}}, X^\delta_{t'_k}, X^\delta_{\tau'_{k+1}}, X^\delta_{\tau_{i+1}}\}$ are in the same

direction. Since, if $X^{\delta}_{\tau'_{k+1}}$ is an extrema point with respect to $X^{\delta}_{t'_k}$ and $X^{\delta}_{\tau_{i+1}}$, then $X^{\delta}_{\tau'_{k+1}} \in E(X^{\delta}, p)_{[0,\tau_{i+1}]}$ which contradicts the assumption. By the same virtue, if $X^{\delta}_{t'_k}$ is an extrema point with respect to $X^{\delta}_{\tau'_{k-1}}$ and $X^{\delta}_{\tau'_{k+1}}$, then $X^{\delta}_{t'_k} \in E(X^{\delta}, p)_{[0,\tau_i]}$ which contradicts the assumption. We assume that the points are increasing, and the decreasing case can be proved by the same virtue. First, we prove that $X^{\delta}_{\tau'_{k-1}} \notin E(X^{\delta}, p)_{[0,\tau_{i+1}]}$.
Let $X^{\delta}_{t'_{k-l}}$ and $X^{\delta}_{t'_{k-l-1}}$ be the closest right and left points to $X^{\delta}_{\tau'_{k-1}}$ from $E(X^{\delta}, p)_{[0,\tau_{i+1}]}$.
Let $\alpha_i = X^{\delta}_{t'_{k-i}} - X^{\delta}_{t'_{k-i-1}}$ for $t'_{k-i} \in E(X^{\delta}, p)_{[0,\tau_{i+1}]}$, where $i = \{0, 1, ..., l-1\}$. By the optimal substructure property, $E(X^{\delta}, p)_{[0,\tau_i]}|_{[\tau'_{k-1},\tau'_{k+1}]} = \{\tau'_{k-1}, \tau'_{k+1}\}$, and we have

$$|X^{\delta}_{\tau'_{k+1}} - X^{\delta}_{\tau'_{k-1}}|^p \geq \sum_{i=0}^{l-1} |\alpha_i|^p + (X^{\delta}_{t'_{k-l}} - X^{\delta}_{\tau'_{k-1}})^p + (X^{\delta}_{\tau'_{k+1}} - X^{\delta}_{t'_k})^p.$$

Since $X^{\delta}_{\tau_{i+1}} - X^{\delta}_{t'_k} > X^{\delta}_{\tau'_{k+1}} - X^{\delta}_{t'_k}$, from Lemma 3.3.2,

$$V_p(\{\tau'_{k-1}, \tau_{i+1}\}, X^{\delta}) > V_p(E(X^{\delta}, p)|_{[\tau'_{k-1},\tau_{i+1}]}, X^{\delta}).$$

Thus $X^{\delta}_{\tau'_{k-1}} \notin E(X^{\delta}, p)_{[0,\tau_{i+1}]}$. In addition, $\{X^{\delta}_{t'_{k-l}}, X^{\delta}_{\tau'_{k-1}}, X^{\delta}_{t'_{k-l+1}}\}$ are in the same direction (increasing by the assumption), since otherwise $X^{\delta}_{\tau'_{k-1}}$ should be included in $E(X^{\delta}, p)_{[0,\tau_{i+1}]}$.
In addition, since $X^{\delta}_{t'_{k-l}} - X^{\delta}_{t'_{k-l-1}} > X^{\delta}_{t'_{k-l}} - X^{\delta}_{\tau'_{k-1}}$, and $X^{\delta}_{\tau_{i+1}} - X^{\delta}_{t'_k} > X^{\delta}_{\tau'_{k+1}} - X^{\delta}_{t'_k}$, and by Lemma 3.3.2,

$$|X^{\delta}_{\tau_{i+1}} - X^{\delta}_{t'_{k-l-1}}|^p \geq (X^{\delta}_{t'_{k-l}} - X^{\delta}_{t'_{k-l-1}})^p + \sum_{i=0}^{l-1} |\alpha_i|^p + (X^{\delta}_{\tau_{i+1}} - X^{\delta}_{t'_k})^p.$$

That is $E(X^{\delta}, p)_{[t'_{k-l-1},\tau_{i+1}]}$ dose not equal to the truncation of $E(X^{\delta}, p)_{[0,\tau_{i+1}]}$ on time interval $[t'_{k-l-1}, \tau_{i+1}]$ which contradicts to the optimal substructure property. □

By Corollary1, the optimal partition satisfies a specific pattern

$$E(X^{\delta}, p)_{[0,\tau_{i+1}]} = E(X^{\delta}, p)_{[0,\tau^*]} \cup \tau_{i+1}$$

where $\tau^* \in E(X^{\delta}, p)_{[0,\tau_i]}$. Therefore the search for optimal partition is essentially equivalent to the search of τ^*, the second last point. We make two observations

about the second last point which allow us to narrow down the spread of the second last point. Recall that $\|X\|_{p,J}$ refers to the p-variation of path $\{X_t\}$ on time interval J.

Observation 3.3.1. *Let τ^* be the second last point to $\{X^\delta_t, t \in [0, \tau_{i+1}]\}$. Then $\tau^* \in E(X^\delta, p)_{[0,\tau_i]}|_{[\tau_k,\tau_i]}$ where τ_k is a point such that $\max_{1\le j\le i} |X^\delta_{\tau_{i+1}} - X^\delta_{\tau_j}|^p = |X^\delta_{\tau_{i+1}} - X^\delta_{\tau_k}|^p$*

Proof. It is obvious $\|X^\delta\|_{p,[0,\tau_l]} \le \|X^\delta\|_{p,[0,\tau_k]}$, $\forall l < k$. Therefore we have

$$\|X^\delta\|_{p,[0,\tau_l]} + |X^\delta_{\tau_{i+1}} - X^\delta_{\tau_l}|^p \le \|X^\delta\|_{p,[0,\tau_k]} + |X^\delta_{\tau_{i+1}} - X^\delta_{\tau_k}|^p$$

Hence t_l for $l < k$, cannot be the last second point. □

Observation 3.3.2. *If there exists $X^\delta_{\tau^i_h}$ and $X^\delta_{\tau^i_{h+1}}$, where $\{\tau^i_h, \tau^i_{h+1}\} \in E(X^\delta, p)_{[0,\tau_i]}$ such that $X^\delta_{\tau^i_h} \le X^\delta_{\tau_{i+1}} \le X^\delta_{\tau^i_{h+1}}$ or $X^\delta_{\tau^i_h} \ge X^\delta_{\tau_{i+1}} \ge X^\delta_{\tau^i_{h+1}}$. Then the last second point of $\{X^\delta_t, t \in [0, \tau_{i+1}]\}$ is in $E(X^\delta, p)_{[0,\tau_i]}|_{[\tau^i_{h+1},\tau_i]}$.*

Proof. It suffices to only consider the case when $X^\delta_{\tau^i_h} \le X^\delta_{\tau_{i+1}} \le X^\delta_{\tau^i_{h+1}}$. Suppose that the second last point is τ^i_k where $k < h$. If $X^\delta_{\tau^i_k} < X^\delta_{\tau^i_h}$,

$$V_p(\{\tau^i_k, \tau_{i+1}\}, X^\delta) < V_p(\{\tau^i_k, \tau^i_{h+1}, \tau_{i+1}\}, X^\delta)$$

and otherwise

$$V_p(\{\tau^i_k, \tau_{i+1}\}, X^\delta) < V_p(\{\tau^i_k, \tau^i_h, \tau_{i+1}\}, X^\delta).$$

In either case τ^i_k cannot be the second last point. □

Algorithm By Theorem 3.3.2, we need only consider the extremal point set $\mathcal{D}_m = \{\tau_0, \tau_1, \tau_2, ..., \tau_m\}$. The algorithm takes an induction on time points, i.e. $E(X^\delta, p)_{[0,\tau_{i+1}]}$ is derived based on $E(X^\delta, p)_{[0,\tau_i]}$. It is an obvious fact that $E(X^\delta, p)_{[0,\tau_1]} = \{\tau_0, \tau_1\}$, and therefore we explain the algorithm given that we know $E(X^\delta, p)_{[0,\tau_2]}$. First, we define an m dimensional vector V with $V(i) = \|X^\delta\|_{p,[0,\tau_i]}$, and thus $V(i) = (\sum_{j=1}^{i} |X^\delta_{\tau_j} - X^\delta_{\tau_{j-1}}|^p)^{\frac{1}{p}}$ for $i = 1, 2$. The algorithm is as follows

1. Given the optimal partition of $\{X^\delta_t, t \in [0, \tau_i]\}$, $E(X^\delta, p)_{[0,\tau_i]}$
2. Find the point τ_l such that $|X^\delta_{\tau_l} - X^\delta_{\tau_{i+1}}| = \max_{\tau \in E(X^\delta, p)_{[0,\tau_i]}} |X^\delta_\tau - X^\delta_{\tau_{i+1}}|$

3. Find the last $\tau_k \in E(X^\delta, p)_{[0,\tau_i]}$ such that $X^\delta_{\tau_k} \leq X^\delta_{\tau_{i+1}} \leq X^\delta_{\tau_{k-1}}$ or $X^\delta_{\tau_k} \geq X^\delta_{\tau_{i+1}} \geq X^\delta_{\tau_{k-1}}$. If we cannot find such τ_k, set $\tau_k = t_0$.

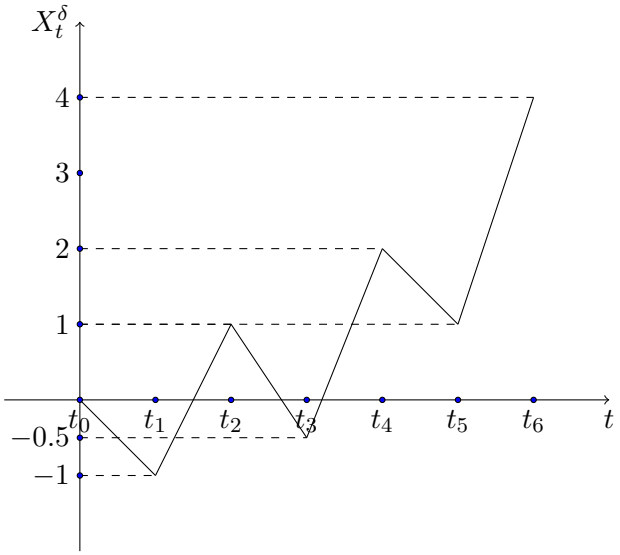
4. Let $\tau_u = max(\tau_l, \tau_k)$

5. Calculate $v_j = V(j)^p + |X^\delta_{\tau_j} - X^\delta_{\tau_{i+1}}|^p$, for $\tau_j \in E(X^\delta, p)_{[0,\tau_i]}|[\tau_u, \tau_i]$

6. Find the τ_m such that $v_m = max\{v_j\}$, and set $V(i+1) = v_m^{\frac{1}{p}}$

7. $E(X^\delta, p)_{[0,\tau_{i+1}]} = E(X^\delta_{\tau_m}, p)_{[0,\tau_m]} \cup \tau_{i+1}$

We would like to make some remark on the algorithm. Step 2 to step 4 are to narrow down the spread of second last points based on Observation 2.3.1 and Observation 2.3.2. In step 5, we find the second last point within the range determined in the previous step. Step 7 is due to Corollary 1.
We would like to illustrate the algorithm by considering the following path



It is obvious that $E(X^\delta, p)_{[t_0,t_4]} = \{t_0, t_1, t_2, t_3, t_4\}$, and we would like to find $E(X^\delta, p)_{[t_0,t_5]}$. We do that by searching for the second last point $t^* \in E(X^\delta, p)_{[t_0,t_4]}$. We apply the two rules to narrow down the spread of the last second point. From step 3, we know that the second last point is t_4, and thus $E(X^\delta, p)_{[t_0,t_5]} = \{t_0, t_1, t_2, t_3, t_4, t_5\}$. For $E(X^\delta, p)_{[t_0,t_6]}$, by applying step 2, we know that the second last point $t^* \in \{t_1, t_2, ..., t_5\}$. We calculate $\|X^\delta\|^p_{p,[0,t_i]} + |X^\delta_{t_6} - X^\delta_{t_i}|^p$ for

$t_i \in E(X^\delta, p)_{[t_0,t_5]}$, and

$$\|X^\delta\|^p_{p,[0,t^*]} + |X^\delta_{t_6} - X^\delta_{t^*}|^p = \max_{i=1,2,\dots,5} \|X^\delta\|^p_{p,[0,t_i]} + |X^\delta_{t_6} - X^\delta_{t_i}|^p.$$

In our case, $t^* = t_3$ for $p = 2$, and $E(X^\delta, p)_{[t_0,t_6]} = \{t_0, t_1, t_2, t_3, t_6\}$.

Complexity Analysis Consider an n-piecewise linear path, and the worst scenario is that, for each time point t_i, we have to check every time point before t_i. Since we start our algorithm from $X^\delta_{t_3}$, the largest number of steps is $\sum_{i=3}^{n-1} i = O(n^2)$. It is worth noticing that $O(n^2)$ is of the same order as the number of edges in the DAG converted by the path, which is consistent with the theorem that the longest path problem consumes linear time $O(|V| + |E|)$ for a DAG $G = (V, E)$ [16]. Numerical results on the running time will be provided in the next section.

3.3.3 Numerical analysis

In this section, we present numerical results on the p-variation calculation. Since p-variation is suggested to be a robust estimator for the diffusion coefficient for multi-scale models [50] for the reason that it can avoid the problem of the optimal sub-sampling, and it can always yield a finite number by changing the value of p, we would also like to present numerical results on the estimation of the diffusion coefficient. The process we consider is fractional Brownian motion $\{B^h_t\}$ where h is the Hurst index, and the process is simulated using fast Fourier transformation [36]. Let the observation interval be $[0, T]$, n be the number of observations, and $\delta = T/n$ the time length between observations. The p-variation should satisfy the condition $ph \geq 1$.

3.3.3.1 Optimal Partition

Fractional Brownian motions are simulated on time interval $[0, 1]$ with $n = 10000$, $n = 5000$, $n = 2500$, $n = 1250$, and $n = 625$. We would like to study the convergent behaviour of the p-variation with respect to n, and we expect that the p-variation and the optimal partition would converge as n increases. We do the experiments for $h = 1/3$, $p = 3$ and $h = 0.7$, $p = 2$ respectively. First, we represent the p-variation for the two cases.

As we can see, the p-variation is larger for the case where $h = \frac{1}{3}$, and it takes

n	$h = \frac{1}{3}$ $p = 3$	$h = 0.7$ $p = 2$
625	2.476 (0.0159)	1.13 (0.0105)
1250	2.643 (0.0360)	1.15 (0.0194)
2500	2.723 (0.0734)	1.17 (0.0455)
5000	2.796 (0.2399)	1.19 (0.1052)
10000	2.887 (0.5448)	1.20 (0.3000)

Table 3.1: The table shows the $p-variation$ of the n-th piecewise-linear paths with different h and p and the computational time in seconds is shown in parenthesis.

more time to compute the p-variation for $h = \frac{1}{3}$. Since fractional Brownian motion oscillates more often as h is smaller, there are more extrema points to consider by the algorithm. We illustrate the variation of computational time with respect to the number of observation by the following plot. As we can see, the case with $h = \frac{1}{3}$ consumes more time to compute the p-variation.

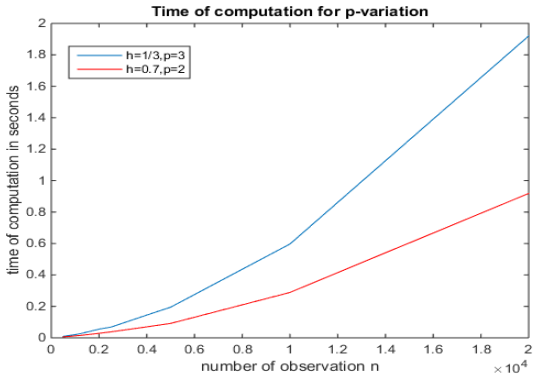


Figure 3.2: The picture shows the computational time in seconds of the n-th piecewise-linear paths with different h and p.

The next set of pictures show the convergence behaviour of the optimal partition. The green line is the original sample path and in order to indicate the convergence behaviour, we show the optimal partition for the linear interpolation of the original sample path. Specifically, we have 10000 observation points for the sample path

on $[0, 1]$, and the linear interpolation of it has 625 observation points. The circle dots and star dots represent the location of the optimal partition. As we can see, the locations of the dots converge quickly for paths of $h = 0.7$, that is the optimal partition for $h = 0.7$ converges faster than the case where $h = \frac{1}{3}$. In addition, the dots are more sparse in the case where $h = 0.7$ than the one where $h = \frac{1}{3}$, because fractional Brownian motion with $h = 0.7$ is less volatile and has less tickles than the fractional Brownian motion with $h = \frac{1}{3}$.

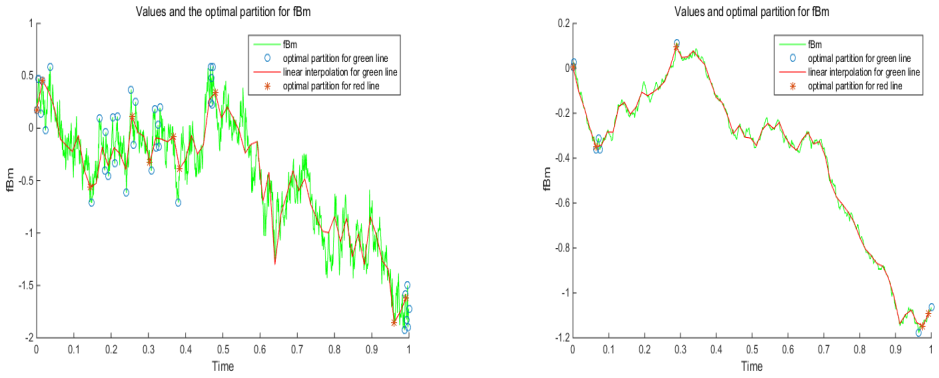


Figure 3.3: The plots show the paths and the optimal partitions. The circle dots and star dots represent the location of the optimal partition. The left hand side is for $h = \frac{1}{3}$, $p = 3$, and the right hand side is for $h = 0.7$, $p = 2$. The red line is the linear interpolation of the sample path (green line)

3.3.3.2 Diffusion coefficient estimation

We perform numerical experiments on the estimation of the diffusion coefficient of a simple process $\{\sigma_0 B_t^h\}$ on time interval $[0, T]$. The true parameter $\sigma_0 = 2$, $T = 10$, and the number of observations $n = 20000$. First, we calculate the mean of $\|B^h\|_{p,[0,T]}$, and we estimate σ_0 as follows,

$$\hat{\sigma} = \frac{\|\sigma_0 B^h\|_{p,[0,T]}}{m(\|B^h\|_{p,[0,T]})},$$

where $m(\|B^h\|_{p,[0,T]})$ is the mean of $\|B^h\|_{p,[0,T]}$. The mean is calculated by taking average of 100 sample paths. The experiment is performed for $h = 0.4$, and $h = 0.7$.

As we can see that $\hat{\sigma}$ performs better when p is closer to $\frac{1}{h}$, that is when we have a more accurate knowledge of the roughness of the path. The estimator performs

$h = 0.4$	$p = 2.6$	$p = 3$
$m(\hat{\sigma})$	1.9997	1.8782
$std(\hat{\sigma})$	0.1692	0.2417
time (second)	2.2868	1.7563

$h = 0.7$	$p = \frac{5}{3}$	$p = \frac{8}{3}$
$m(\hat{\sigma})$	2.0185	1.9912
$std(\hat{\sigma})$	0.4180	0.6732
time(second)	1.2018	1.1488

Figure 3.4: The table shows the results of the estimation of the diffusion coefficient. The true parameter $\sigma_0 = 2$, $T = 10$, and the number of observations $n = 20000$.The mean $m(.)$ and standard deviation $std(.)$ are calculated by taking 100 Monte Carlo simulations, and time refers to average time for computing one path in seconds

better with smaller h, since paths with smaller h reflects more information on volatility. When $h > \frac{1}{2}$, the path is of long memory which makes the estimation of diffusion coefficient harder.

3.4 Multidimensional case

We consider the algorithm for multidimensional paths in this section. The multidimensional piecewise linear path can be converted to a DAG by the same way of the one-dimensional case does. Consider a d- dimensional piecewise linear path $\{X_{t_0}, \dots, X_{t_n}\}$, and for each time point t_i, $X_{t_i} = (X^1_{t_i}, \dots, X^d_{t_i})$. The corresponding DAG can be constructed by taking each point t_i as a vertex and adding directed edges (t_i, t_j) if $i < j$. However, the weights are calculated differently from the one-dimensional case, and for instance, the weight for edge (t_i, t_j), is

$$w(t_i t_j) = (\sum_{l=1}^{d} |X^{\delta,(l)}_{t_i,t_j}|^r)^{\frac{p}{r}},$$

by taking the vector norm the l^r norm. Therefore, the longest path problem has the optimal substructure for the multidimensional path, and the algorithm only requires a linear time of calculation. The worst scenario requires $O(n^2)$ which is the same as the one-dimensional case where n is the number of piecewise linear interpolation. Another remark we want to make is that the rules applying to the one dimensional path to narrow down the range of the candidate points (such as the extremal point set, and the two rules for calculating the spread of second last points) do not apply to the multidimensional path. Since the first problem we encounter is the definition of the extremal point, which does not apply to the multidimensional paths. Thus, our search starts from the original partition $\mathcal{D}_0$.

3.4.1 Algorithm

We present the algorithm for multi-dimensional paths which is similar to the one dimensional case. The DAG of the multidimensional path and the one dimensional path is the same besides the weight, since the weight for multidimensional paths requires calculating the vector norms. Thus the algorithm for multidimensional paths is equivalent to the search of the second last point, and the candidate pool is a subset of the optimal partition by the optimal substructure property.
Let $\{X_t^\delta, t \in [0,T]\}$ be a d-dimensional piecewise linear path on homogeneous time point set

$$\mathcal{D}_0 = \{t_0, t_1, ..., t_n\} = \{0, \delta, 2\delta..., T\},$$

which we call the original partition. For t_i, we have $X_{t_i}^\delta \in (\mathbb{R}^d, \|.\|)$, where $\|.\|$ refers to the vector norm. We starts from the original partition $\mathcal{D}_0$, and since the optimal partition $E(X^\delta, p)_{[0,t_1]} = \{t_0, t_1\}$, we starts from t_2. Suppose we are given that the optimal partition of $\{X_t^\delta, t \in [0,t_i]\}$, and we present the the algorithm to find the optimal partition for $\{X_t^\delta, t \in [0,t_{i+1}]\}$. First, we define an n dimensional vector V which stores the p-variation of the path from t_0 to t_i, and thus $V(i) = \|X^\delta\|_{p,[t_0,t_i]}$. For instance, $V(1) = \|X_{t_1}^\delta - X_{t_0}^\delta\|$. The algorithm is as follows

1. Calculate $a_j = V(j)^p + \|X_{t_j}^\delta - X_{t_{i+1}}^\delta\|^p$, for $t_j \in E(X^\delta, p)_{[0,t_i]}$
2. Find the t_m such that $a_m = \max_{j \in E(X^\delta,p)_{[t_0,t_i]}} a_j$, and set $V(i+1) = a_m^{\frac{1}{p}}$.
3. $E(X^\delta, p)_{[0,t_{i+1}]} = E(X^\delta, p)_{[0,t_m]} \cup t_{i+1}$.

3.4.2 Numerical analysis

The numerical experiments are performed for a 2-dimensional fraction Brownian motion. The calculation of p-variation and estimation of diffusion coefficient problem are investigated. We take the l^2 norm as the vector norm for the multi-dimensional paths.

3.4.2.1 Optimal partition

First, we study the convergence behaviour of the optimal partition and the p-variation of a two-dimensional fractional Brownian motion with respect to n, the number of observations. We expect that the p-variation and the optimal partition to converge as n increases. The setting of the numerical experiments are as follows.

Fractional Brownian motions are simulated on fixed time interval $[0,1]$ with $n = 10000$, $n = 5000$, $n = 2500$, $n = 1250$, and $n = 625$. We do the experiments for $h = 0.4$, with $p = 2.6$ and $p = 3$, and $h = 0.7$ with $p = \frac{5}{3}$ and $p = \frac{8}{3}$. We do experiments on different p in order to see the effect of p on the optimal partition. We study the behaviour of the p-variation first.

n	$p = 2.6$	$p = 3$
625	2.8660	2.7515
1250	3.0111	2.8669
2500	3.0756	2.8815
5000	3.0949	2.8975
10000	3.1613	2.9156

n	$p = \frac{5}{3}$	$p = \frac{8}{3}$
625	1.9476	1.8276
1250	1.9593	1.8344
2500	1.9672	1.8382
5000	1.9707	1.8386
10000	1.9712	1.8403

Table 3.2: The tables show the p-variations of two n-piecewise-linear fractional Brownian motions on $[0,1]$. The left-hand side table is when $h = 0.4$, and the right-hand side table is when $h = 0.7$.

From the table, we can see that the p-variation is larger for the case where h is smaller, since fractional Brownian motion is rougher when h is smaller. For a fixed Hurst index h, p-variation is smaller when p is larger, and we expect that the corresponding optimal partition would converge faster than the optimal partition with smaller p.

n	$p = 2.6$	$p = 3$
625	0.0056	0.0054
1250	0.0151	0.0130
2500	0.0420	0.0368
5000	0.1507	0.1171
10000	0.5158	0.4465

n	$p = \frac{5}{3}$	$p = \frac{8}{3}$
625	0.0047	0.0041
1250	0.0119	0.0098
2500	0.0356	0.0296
5000	0.1216	0.0877
10000	0.4715	0.3997

Table 3.3: The tables show the computation time in seconds for the p-variations of two n-piecewise-linear fractional Brownian motions on $[0,1]$. The left-hand side table is when $h = 0.4$, and the right-hand side table is when $h = 0.7$.

The above tables show the calculation time with unit in second, and as we can see that the algorithm consumes less time for paths with larger h, since such paths are smoother and the optimal partitions are more sparse than the paths with smaller h. When fixing h, the computation time is less when p is larger.
The next set of pictures show the convergence behaviour of the optimal partition. The green line is the original sample path and in order to indicate the convergence behaviour, we show the optimal partition for the linear interpolation of the original

sample path. Specifically, we have 10000 observation points for the sample path on $[0, 1]$, and the linear interpolation of it has 625 observation points. The circle dots and star dots represent the location of the optimal partition. As we can see, the optimal partition for $h = 0.7$ converges faster than case where $h = 0.4$, since the locations of the dots converge quickly for paths of $h = 0.7$. As we can see, for both pictures, when h is fixed, the higher dots representing the location of the optimal partition of larger p is more sparse and converge faster than the lower ones which is the locations of the optimal partition of larger p.

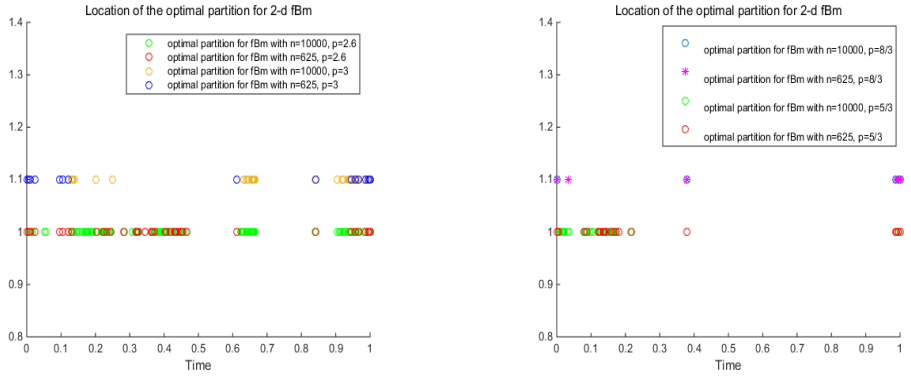


Figure 3.5: The plots show the locations of the optimal partitions. The circle dots and star dots represent the location of the optimal partition. The left hand side is for $h = 0.4$, $p = 2.6$, and $p = 3$; the right hand side is for $h = 0.7$, $p = \frac{5}{3}$ and $p = \frac{8}{3}$.

3.4.2.2 Diffusion coefficient estimation

We perform numerical experiments on the estimation of the diffusion coefficient of a simple process $\{\sigma_0 B_t^h\}$ on time interval $[0, T]$. The true parameter $\sigma_0 = 2$, $T = 10$, and the number of observations $n = 20000$. First, we calculate the mean of $\|B^h\|_{p,[0,T]}$, and we estimate σ_0 as follows,

$$\hat{\sigma} = \frac{\|\sigma_0 B^h\|_{p,[0,T]}}{m(\|B^h\|_{p,[0,T]})},$$

where $m(\|B^h\|_{p,[0,T]})$ is the mean of $\|B^h\|_{p,[0,T]}$. The mean is calculated by taking average of 100 sample paths. The experiment is performed for $h = 0.4$, and $h = 0.7$.

$h = 0.4$	$p = 2.6$	$p = 3$	$h = 0.7$	$p = \frac{5}{3}$	$p = \frac{8}{3}$
$m(\hat{\sigma})$	1.9999	1.9874	$m(\hat{\sigma})$	2.0054	1.9967
$std(\hat{\sigma})$	0.0086	0.0501	$std(\hat{\sigma})$	0.3699	0.5868
time (second)	2.5009	1.7194	time (second)	1.9846	1.4673

Table 3.4: The table shows the results of the estimation of the diffusion coefficient. The true parameter $\sigma_0 = 2$, $T = 10$, and the number of observations $n = 20000$.The mean $m(.)$ and standard deviation $std(.)$ are calculated by taking 100 Monte Carlo simulations, and time refers to average time for computing one path in seconds

From the table, we can see that the case with smaller h performs better than the case with a larger h. As paths with smaller h reflects the volatility more strongly. However, the calculation time for the case with larger h is less. As is shown in the previous section, paths with larger h consumes less time when computing the p-variation and optimal partition. In addition, the same as the one-dimensional case, $\hat{\sigma}$ performs better when p is closer to $\frac{1}{h}$, that is when we have a more accurate knowledge of the roughness of the path.

3.5 Conclusion

In the first project, we construct an efficient algorithm for finding the optimal partition for a piecewise linear path. The algorithm transforms the path into a DAG and it possesses optimal substructure which guarantees the algorithm consumes linear time. Numerical experiments are performed for the calculation of p-variation. Since it is suggests by A.Papavasiliou [50] that the p-variation is a robust estimator for the diffusion coefficient for the multiscale model, numerical experiments are also performed on the estimation of diffusion coefficient. The estimator is close to the true parameter, but there are many theoretical problems required to be solved for the estimator to be useful, for example, whether the mean of the p-variation of a path will converge as the number of observation goes to infinity, and whether the estimator is consistent.

Chapter 4

Statistical inference for discretely observed processes

This chapter studies the problem of constructing the approximate MLEs for discretely observed rough stochastic differential equations (RSDEs). For most of the cases in practice, the data comes in discrete form, but there are not much studies on the problem of parameter estimations under the discrete observation settings. The common practice is to construct the estimators from the continuously observed case, and then descretize the estimator for it to be applicable for the discretely observed case. In addition, most of the studies focus on the diffusion processes, and the problem for processes beyond that category are studied less. Motivated by rough path theory, the method proposed by A.Papavasiliou and K.Taylor[52] is able to cover a wider class of processes. Thus, we focus ourselves on the fractional processes that is the stochastic processes driven by fractional Brownian motion. The method projects the continuous process to a tractable process which follows a sequence of piecewise smooth ordinary differential equations (ODEs). The ODEs are constructed by discretising the driving force on the set of the discrete observation time points, and the likelihood function for the tractable process is constructed which is taken to be the approximate likelihood function. The method has several advantages. First, it is easy to apply in a sense that the step is straightforward without involvement with advanced stochastic calculus. In addition, the original model is transformed to a series of ODEs which usually have better analytical properties. Since the discritized driving force can be parametrised by a finite dimensional random variable, the probability density function is tractable. For instance, the approximate likelihood function of a Gaussian process is multidimensional normal.

Second, the method can be applied to a more general class of processes. The motivation comes from the rough path theory that allows the solutions of the ODEs to converge to the continuous p-rough path in p-variation metric for geometric p-rough paths. Thus, p need not be constrained within the range of smaller than two. Whereas most methods [32], [33], only apply to the paths with $p < 2$, this method has wider range of applications.
We study the method by applying it to the fractional Ornstein-Unlenbeck (O-U) process, which possesses analytical solution. Therefore, it is possible to study the properties of the MLEs analytically. In addition, most of the methods for the parameter estimation problem for fractional processes study the fractional O-U process, and we can compare our method to other methods using the case of fractional O-U process.
The chapter can be divided into two parts. The first part studies the topic in general. First, we provide background knowledge on fractional calculus. Then, some literature review on the topic of parameter estimation for fractional processes is presented. Afterwards, the convergence property of the log-likelihood function is investigated. Specifically, by assuming that the paths converge in p-variation metric, we would like to know whether the likelihood function would convergence. This helps us to understand the consistency property of the approximate MEs. In the second part, we apply the method to the one-dimensional fractional O-U process for both the drift and diffusion coefficient.

4.1 Preliminaries on fractional calculus

The study of the fractional processes is closely related to fractional calculus, since the fractional processes are usually represented as fractional integrals which we will explain later in this section. Thus, it is necessary to present some preliminaries on fractional calculus. because we focus on the fractional processes in our project. The material on the preliminaries of the fractional calculus can be found in [46] (Chapter 1.1 (p1-p7)). The fractional integral is from the Riemann-Liouville integral. Consider a Lebesgue measurable function $f : [a, b] \to \mathbb{R}$. Let $0 < \alpha < 1$, and the Riemann-Liouville integral is defined as,

$$(I^{\alpha}_{a^+} f)(x) = \frac{1}{\Gamma(\alpha)} \int_a^x f(s)(x-s)^{\alpha-1} ds,$$

and

$$(I_{b^-}^{\alpha} f)(x) = \frac{1}{\Gamma(\alpha)} \int_x^b f(s)(s-x)^{\alpha-1} ds.$$

For $f : \mathbb{R} \to \mathbb{R}$, we have

$$(I_{-}^{\alpha} f)(x) = \frac{1}{\Gamma(\alpha)} \int_x^{\infty} f(s)(s-x)^{\alpha-1} ds.$$

We say a function is in the domain of the operator, $f \in D(I_{-}^{\alpha})$, if the integration converges almost everywhere. We introduce the following theorem which will be useful in the later proof,

Theorem 4.1.1. *(Chapter 1.1, p1, Theorem 1.1.1) Let $1 \leq p,\ q < \infty$, and $0 < \alpha < 1$. The operator I_{-}^{α} is bounded from $L_p(\mathbb{R})$ to $L_q(\mathbb{R})$ if and only if $1 < p < \frac{1}{\alpha}$, and $q = \frac{p}{1-\alpha p}$.*

Define $f \in I_{-}^{\alpha}(L_p(\mathbb{R}))$, if $f = I_{-}^{\alpha}(g)$ almost everywhere, and $g \in L_p(\mathbb{R})$. In order to find function g, we need to find the inverse of the operator I_{-}^{α}. For $0 < \alpha < 1$, and $f \in I_{-}^{\alpha}(L_p(\mathbb{R}))$ with $p \geq 1$, the inverse of the operator I_{-}^{α} which is also called the Riemann-Liouville fractional derivative is denoted as follows, (p3)

$$(I_{-}^{-\alpha} f)(x) = (D_{-}^{\alpha} f)(x) = \frac{-1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_{-\infty}^x f(s)(x-s)^{-\alpha} ds.$$

The Riemann-Liouville fractional derivative can also be defined for $f : [a, b] \to \mathbb{R}$ where $f \in I_{a^+}^{\alpha}(L_p([a, b]))$ or $f \in I_{b^-}^{\alpha}(L_p([a, b]))$ as follows (p3),

$$(I_{a^+}^{-\alpha} f)(x) = (D_{a^+}^{\alpha} f)(x) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_a^x f(s)(x-s)^{-\alpha} ds,$$

and

$$(I_{b^-}^{-\alpha} f)(x) = (D_{b^-}^{\alpha} f)(x) = \frac{-1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_x^b f(s)(x-s)^{-\alpha} ds.$$

They admit Weyl representation of fractional derivatives (p3-p4):

$$(D^{\alpha}_{a^+}f)(x) = \frac{1}{\Gamma(1-\alpha)}\left(f(x)(x-a)^{-\alpha} + \alpha\int_a^x (f(x)-f(s))(x-s)^{-\alpha-1}ds\right)\mathbb{1}_{(a,b)}(x)$$

$$(D^{\alpha}_{b^-}f)(x) = \frac{1}{\Gamma(1-\alpha)}\left(f(x)(b-x)^{-\alpha} + \alpha\int_x^b (f(x)-f(s))(s-x)^{-\alpha-1}ds\right)\mathbb{1}_{(a,b)}(x)$$

The equality is in pointwise sense for $p = 1$, and in $L_p([a,b])$ sense for $p > 1$. The following operator is defined in [46] (Chapter 1.3, p10, equation 1.3.3),

$$M^h_- f = \begin{cases} C(h)I^{\alpha}_-(f) & h \in (0,1)\setminus\frac{1}{2} \\ f & h = \frac{1}{2}, \end{cases}$$

with $\alpha = h - \frac{1}{2}$, and $C(h) = \Gamma(h+\frac{1}{2})\frac{(2hsin\pi h\Gamma(2h))^{\frac{1}{2}}}{\Gamma(h+1)}$. If $M^h_-(f)(s) \in L_2$, we say that $f \in L^h_2$., and he Wiener integral with respect to fractional Brownian motion is defined as follows (Chapter 1.6, p16, Definition 1.6),

Definition 8. *For $f \in L^h_2$, the integral with respect to fractional Brownian motion is*

$$\int_{\mathbb{R}} f(s)dB^h_s := \int_{\mathbb{R}} M^h_-(f)(s)dW_s.$$

We introduce the following theorem,

Theorem 4.1.2. *(Chapter 1.6, p16, Lemma 1.6.2) For $h \in (0,1)$, the linear span $\{M^h_-(\mathbf{1}_{s,t}), s,t \in \mathbb{R}\}$ is dense in L_2.*

Thus the Wiener integral can be defined as

$$\int_{\mathbb{R}} f(s)dB^h_s = \lim_{m\to\infty}\int_{\mathbb{R}} M^h_-(f_m)(s)dW_s,$$

where $f_m(x) = \begin{cases} f(\frac{i}{m}) & x \in [\frac{i}{m}, \frac{i+1}{m}) \\ 0 & otherwise \end{cases}$ is the step function of order m, and the

limit is in $L_2(\Omega)$. In addition, we obtain that

$$W_t = \frac{1}{C(h)} \int_{\mathbb{R}} (I_-^{-\alpha} \mathbf{1}_{(0,t)}) dB_s^h. \tag{4.1}$$

4.2 Literature Review

The literature review focuses on the topic of parameter estimation for fractional processes, and statistical application of rough path theory.

4.2.1 Parameter estimation for the drift coefficient

First, we consider the parameter estimation for the drift coefficient. Most researches adopt two types of estimators for the drift coefficient, maximum likelihood estimator (MLE) and least square estimator (LSE). First, we introduce the MLE. In [64], C.Tudor and F.Viens construct the MLE for the drift coefficient of a fractional differential equation given by

$$dY_t = \lambda b(Y_t)dt + dB_t^h, \quad Y_0 = 0 \tag{4.2}$$

where $\lambda \in \Lambda$ is the drift coefficient with Λ a compact interval, $b(Y_t)$ is a Lipschitz continuous function which has a specific form (we know the exact function of $b(Y_t)$), and $\{B_t^h, t \in [0,T]\}$ is a fractional Brownian motion in $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with $h \in (0,1)$. By assuming Lipschitz condition on b, they proved that $\{Y_t\}$ admits the following Radon-Nikodym derivative,

$$\frac{d\tilde{P}}{dP} = exp\left(-\lambda \int_0^t K_h^{-1}(\int_0^{\cdot} b(Y_u)du)(s)dW_s - \frac{1}{2}\lambda^2 \int_0^t (K_h^{-1}(\int_0^{\cdot} b(Y_u)du)(s))^2 ds \right), \tag{4.3}$$

where $\{W_s\}$ is a P-Brownian motion, and K_h^{-1} is, for $h > \frac{1}{2}$,

$$(K_h^{-1} f)_s = s^{h-\frac{1}{2}} D_{0^+}^{h-\frac{1}{2}} (s^{\frac{1}{2}-h} f'(s))(s),$$

and for $h < \frac{1}{2}$,

$$(K_h^{-1} f)_s = s^{h-\frac{1}{2}} I_{0^+}^{\frac{1}{2}-h} (s^{\frac{1}{2}-h} f'(s))(s).$$

Recall that for $f : [0, T] \to \mathbb{R}$,

$$(D^{\alpha}_{0^+} f)(x) = \frac{1}{\Gamma(1-\alpha)} \left(f(x)x^{-\alpha} + \alpha \int_0^x (f(x) - f(s))(x-s)^{-\alpha-1} ds \right) \mathbb{1}_{(0,T)}(x),$$

and

$$(I^{\alpha}_{0^+} f)(x) = \frac{1}{\Gamma(\alpha)} \int_0^x f(s)(x-s)^{\alpha-1} ds.$$

The resulting MLE is as follows, and they proved strong consistency for $h \in [0, 1]$.

$$\hat{\lambda}(y)_T = \frac{\int_0^T K_h^{-1}(\int_0^{\cdot} b(y_u)du)(s) d(\int_0^s K_h^{-1}(s,u)dY_u)}{\int_0^T (K_h^{-1} \int_0^{\cdot} b(Y_u)du)(s)^2 ds}. \tag{4.4}$$

They made a brief discussion on the discrete case where the observation time points are integers, i.e. $[0, 1, 2, ..., n]$. The estimator is obtained by discretising the estimator (4.4) obtained in the continuous case

$$\hat{\lambda}(y)_n = \frac{\sum_{m=0}^{n} Q_m(Z_{m+1} - Z_m)}{\sum_{m=0}^{n} |Q_m|^2}, \tag{4.5}$$

where $Q_m = c(H)n^{H-\frac{1}{2}} \sum_{j=0}^{m-1} (m-j)^{-H-\frac{1}{2}} j^{\frac{1}{2}-H} y_j$, with $c(H)$ a constant depending on H and $\{y_i\}$ the observations, and $Z_m = \sum_{j=0}^{m-1} K^{-1}(m, j)(y_{j+1} - y_j)$, with $K^{-1}(m, j)$ being the inverse of the Molchan-Colosov kernel on integer time points. A list of assumptions are given for the proof of the consistency of the estimators, but no conclusive result is given.

Since it is difficult to derive the likelihood function for fractional processes, some researchers turn to least square estimator(LSE) to avoid the difficulty of constructing the likelihood. In [32], Hu and Nualart consider the following fractional O-U process

$$dY_t = -\lambda Y_t dt + \sigma dB^h_t, \quad Y_0 = 0, \tag{4.6}$$

where $\{B_t^h, t \in [0,T]\}$ is a fractional Brownian motion with $h \geq \frac{1}{2}$, and $\lambda, \sigma > 0$ where σ is known. They proved strong consistency for the following estimator

$$\tilde{\lambda}_T = -\frac{\int_0^T Y_t \delta Y_t}{\int_0^T Y_t^2 dt} = \lambda + \sigma \frac{\int_0^T Y_t \delta B_t^h}{\int_0^T Y_t^2 dt}, \tag{4.7}$$

where $\int .\delta Y$ refers to the Skorohod integral [49]. The numerator $\int_0^T Y_t \delta B_t^H$ in (4.7) is pathwise integrable by Young's theorem, but the expectation is not zero in pathwise sense. Therefore the integral is defined in Skorohod sense, and since the skorohod integral includes a compensator in addition to the pathwise integral, $\tilde{\lambda}_T$ is not biased. Although (4.7) is unbiased and strong consistent, it is found not applicable because by the calculation of the parameter,

$$\tilde{\lambda}(y)_T = -\frac{Y_T^2}{2\int_0^T Y_t^2 dt} + \sigma^2 \frac{\alpha_h \int_0^T \int_0^t e^{-\lambda(t-s)}(t-s)^{2h-2} ds dt}{\int_0^T Y_t^2 dt},$$

we can see that the estimator contains the unknown parameter λ. Though the method is not practical, it has several extensions. R.Belfadli et al [7] consider the fractional O-U process where λ is negative, the non-ergodic case, and the least squared error estimator is constructed as (4.7). However, the stochastic integral is defined in pathwise sense, and strong consistency result is obtained for $h > \frac{1}{2}$. By defining the stochastic integral in a pathwise sense, the estimator is applicable, since it does not involve the unknown information λ as in the skorohod integral.
By discretizing the estimator (4.7), an approximate estimator for the drift coefficient for the discretely observed case is constructed by K.Es-Sebaiy as follows [24],

$$\tilde{\lambda}(y)_n = \frac{\sum_{m=1}^{n} \int_{(m-1)\Delta}^{m\Delta} y_{m-1} \delta y_s}{\sum_{m=1}^{n} y_{m-1}^2},$$

where Δ is the time between observations, n is the number of observations, and the integral is an Skorohod integral. K.Es-Sebaiy prove that $\tilde{\lambda}(y)_n$ converges in probability when $\Delta \to 0$ and $n\Delta \to \infty$. However, the estimator is not applicable due to the unobservable part of Skorohod integral.
A third estimator which only applies to fractional O-U process is constructed by

Hu and Nualart,

$$\bar{\lambda}(y)_T = \left(\frac{1}{\sigma^2 h T \Gamma(2h)} \int_0^T y_s^2 ds \right)^{\frac{1}{-2h}}. \tag{4.8}$$

The corresponding discrete version of (4.8) is discussed in [33], where Hu and Song prove that the following estimator

$$\bar{\lambda}(y)_n = \left(\frac{1}{\sigma^2 h n \Gamma(2h)} \sum_{m=1}^{n} y_m^2 \right)^{\frac{1}{-2h}} \tag{4.9}$$

is strongly consistent as $n \to \infty$ without any requirement on δ.
We make compare the three types of estimators. MLE (4.4) has the widest range of application. The only assumption for the stochastic differential equation is that the function in the drifting part is Lipschitz, and it can be applied to any h-fBM with $h \in (0, 1)$ whereas other estimators can only be applied to h-fBM with $h > \frac{1}{2}$. However, the MLE (4.4) only applies to continuously observed case and no conclusive result on the discrete case can be drawn from the the paper [64]. The LSE (4.7) is proved to be weakly consistent for discretely observed case but the estimator is not applicable since it contains the parameter which we want to estimate. However, the LSE can be applied to non-ergodic case for fractional O-U process with $h > \frac{1}{2}$. The last estimator (4.8) is easy to calculate and strongly consistent. However, (4.8) has limited application, since it can only be applied to cases with $h \geq \frac{1}{2}$ and known diffusion coefficient. Besides, the method cannot be generalised to other stochastic model, though it is an efficient estimator for fractional O-U process.

4.2.2 Parameter estimation for the diffusion coefficient

For the estimation of diffusion coefficient, the basic idea is to use the variation of the observations as the estimator, since the information on diffusion coefficient is reflected in the variation of the path. Quadratic variation has been a popular estimator for the diffusion coefficient, for instance [30], [5], and [2]. However, quadratic variation has its limitation, and for example, as it is suggested by [26], [34], the quadratic variation of the stock price will diverge if data is obtained on a very fine grid. Thus generalised quadratic variation is proposed as an alternative estimator for the diffusion coefficient, for instance, A.Brouste [9] proposed the estimators for Hurst and volatility parameters using the technique of generalised quadratic variation. Let $\mathbf{a} = (a_0, ..., a_K)$ be a discrete filter of length $K + 1$ with $K \in \mathbb{N}$, and of

order $L \geq 1$, and $K \geq L$, i.e.

$$\sum_{k}^{K} a_k k^l = 0, \quad \sum_{k}^{K} a_k k^L \neq 0,$$

where $0 \leq l < L$. For such a filter $\mathbf{a}$, let $V_{N,\mathbf{a}} = \sum_{i=0}^{N-K} (\sum_{k=0}^{K} a_k X_{i+k})^2$. Let the filter be normalised as follows,

$$\sum_{k=0}^{K} (-1)^{1-k} a_k = 1,$$

and the dilated filter $\mathbf{a}^2$ associated with $\mathbf{a}$ is as follows, for $0 \leq k \leq 2K$,

$$a_k^2 = \begin{cases} a(2k) & if k = 2k' \\ 0 & otherwise \end{cases}.$$

Then the estimators are as follows

$$\hat{h}_N = \frac{1}{2} log(\frac{V_{N,a_2}}{V_{N,a}}), \tag{4.10}$$

and

$$\hat{\sigma}_N = (2 - \frac{V_{N,a}}{\sum_{k,l} a_k a_l |k-l|^{2\hat{h}_N} \Delta^{\hat{h}_N}})^{\frac{1}{2}}, \tag{4.11}$$

Strong consistency results are derived and numerical analysis on the method is presented in [9]. As volatility is an important parameter for financial model, there are many studies on the volatility estimation for financial model. In [15], A.Chronopoulou models the volatility of stock price $\{Y_t\}$ as follows,

$$dY_t = \alpha(m - Y_t)dt + \beta dB_t^h,$$

where B_t^h is an h-fractional Brownian motion. The variogram defined as $V_j^N = \frac{1}{N} \sum_{n=1}^{N} (L_{n+j} - L_n)^2$ can be considered as the expected quadratic variation of lag of j, where L_n is the implied volatility defined as

$$L_n = log(|\frac{2(X_n - X_{n-1})}{\sqrt{\Delta}(X_n + X_{n-1})}|),$$

and X_t is the price of the stock at time t. By deriving the approximate theoretical value of variogram of different lag j, a system of equations is constructed and the estimators are derived by solving the system of equations. The method shares the same virtue of moment matching method, but instead of matching moments, they match the variogram. The above two methods rely on the notion of variance of a process. Other methods to construct the diffusion coefficient include using the number of crossings [38], and maximizing the spectral likelihood function[8]. However, they are constrained to the continuous fractional process with $h > \frac{1}{2}$.

4.2.3 Rough path and statistics

Since our method is motivated by rough path theory, we would like to present a brief literature review on the application to the statistical inference problem by rough path theory. The first paper on the application of rough path theory to statistical inference goes back to 2011 [51], in which the expected signature matching method is proposed within the framework of rough differential equation. For a d-dimensional path $\{Y_t\}$, the m-th level of iterated integral is defined as

$$Y^{(m)}_{[0,T]} = \sum_{w_i \in \sigma} \int ... \int_{0<t_1<t_2...<t_m<T} dY^{i_1}_{t_1} ... dY^{i_m}_{t_1},$$

where $w_i = \{i_1, ..., i_m\}$ is a word over $\{1, 2, ..., d\}$, and σ is the set of all the possible words. The estimators are the solutions of the equation between the experimental expectation of $Y^{(m)}_{[0,T]}$ and the theoretical value.
Another pioneer paper discussing the application of rough path to the statistical problem is by D.Levin, T.Lyons and H.Ni [39]. They discussed the possibility and advantage of using the signature of a path as the feature set for statistical problems, and illustrate the argument by applying the idea to the regression problem. The expected signature model is defined as follows,

Definition 9. *(Definition 3.3, p14, [39]) Let X and Y be two stochastic processes taking values in E and W respectively. Suppose that the signatures of X and Y denoted by $\mathbf{X}$ and $\mathbf{Y}$ are well defined a.s. Assume that*

$$\mathbf{Y} = L(\mathbf{X}) + \epsilon,$$

where $\mathbb{E}[\epsilon|\mathbf{X}] = 0$, and L is a linear functional mapping $S(E)$ to $S(W)$.

Recall that $S(E)$ refers to the signature space of E. The advantage of the above model is that it can avoid the curse of dimensionality due to high frequency of sampling. Because the dimension of the signature only depends on the dimension of the path and the degree of the truncated signature. Secondly, the signature can store the information in a structural way which is insensitive to the time sampling dimension, since the signature captures the essential information of the path on a time segment and does not change by the parametrization of time.
Numeric results on AR, and polynomial autoregressive models, which suggest that the R^2 of the prediction is better as the level of the iterated integral is higher. In comparison to the AR method, the results suggest that the expected signature method can be applied to a more wider types of data at an equal or better accuracy of prediction. When comparing to Gaussian processes method, the expected signature method achieves similar accuracy at a much smaller computational cost. Diehl, Friz and Mai propose a pathwise stable MLE for the drift coefficients of multidimensional diffusion processes [20] by lifting the classical MLE up in the rough path setting. They consider a d-dimensional differential equation,

$$dY_t = h(Y_t)Adt + \Sigma(Y_t)dB_t,$$

where $\{B_t, t \in [0,T]\}$ is a d-dimensional Brownian motion, $A \in V$, some finite-dimensional vector space, and sufficient regular $h : \mathbb{R}^d \to L(V, \mathbb{R}^d)$, and $\Sigma : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R}^d)$. They derive the MLE for A by lifting the MLE constructed by Girsanov's theorem to rough path space, that is

$$\hat{A}_T(\mathbf{Y}) = (\int_{t=0}^{T} h(Y_t)^* C^{-1}(Y_t) h(Y_t) dt)^{-1} \int_{t=0}^{T} h(Y_t) C^{-1}(Y_t) d\mathbf{Y}_t, \tag{4.12}$$

where $\mathbf{Y}$ is a $2+\epsilon$ rough path with $\epsilon > 0$, $C(Y_t) = \Sigma(Y_t)\Sigma(Y_t)^T$, and the integral is in Stratonovich sense. They proved that the estimator (4.12) is pathwise stable in rough path topology. That is if $d_p(\mathbf{Y}, \mathbf{X}) << 1$, we have $|\hat{A}_T(\mathbf{X}) - \hat{A}_T(\mathbf{Y})| << 1$. They investigate the case of mis-specification of error as well. Consider the case where the actual driving force is a fractional Brownian motion, $\{B_t^h\}$ with h close to $\frac{1}{2}$. Let $\mathbf{Y}^h$ be the solution of the rough differential equation driven by $\{B_t^h\}$. Since $d_p(\mathbf{Y^h}, \mathbf{Y})_{[0,T]} << 1$ [27], $|\hat{A}_T(\mathbf{Y}^h) - \hat{A}_T(\mathbf{Y})| << 1$. That implies if we move the estimator to a rough path topology, we can apply the estimator to fractional processes and at the same time the error is controlled. The result suggests a neces-

sity to move to rough path framework, since more robust results can be obtained. Another pathwise stable estimator obtained by applying rough path theory is constructed by Qian and Xu [56]. They consider the multi-dimensional fractional O-U process with $h \in (\frac{1}{3}, \frac{1}{2})$

$$dX_t = -\Lambda X_t dt + \sigma dB_t^h, \qquad X_0 = x_0,$$

where $\{B_t^h\}$ is a d-dimensional fractional Brownian motion, and Λ a positive symmetric $d \times d$ matrix. They construct the LES for the drift matrix Λ in rough path sense as follows,

$$\hat{\Lambda}_t = -(\int_0^t X_s \otimes X_s ds)^{-1}(\int_0^t X_s \otimes d\mathbf{X}), \tag{4.13}$$

where the integral is defined in Ito's sense. They proved the strong consistency result for the estimator for the continuously observed case. They also construct the estimator for the discretely observed case by discretizing the estimator (4.13),

$$\hat{\Lambda}_n = -(\sum_{l=0}^{n}(X_{l\delta} \otimes X_{l\delta})\delta)^{-1}(\sum_{l=0}^{n-1} X_{l\delta}X_{l\delta,(l+1)\delta} + \mathbb{X}_{l\delta,(l+1)\delta}),$$

where $\otimes$ is the tensor product, $\mathbb{X}_{s,t}$ is the second level iterated integral of $\{X_t\}$, and $\delta := \frac{T}{n}$ is the time between observations. They proved that the estimator is strong consistent if $n\delta \to \infty$, $\delta \to 0$, and $n\delta^q \to 0$ for some $p \in (1, \frac{1+h+\beta}{1+\beta})$, and $0 < \beta < 1$.

4.3 Main idea and the setting of the problem

We introduce the setting and the main idea of the method proposed by A.Papavasiliou and T.Kasia [52] in this section. We consider the following differential equation,

$$dY_t = f(Y_t;\theta)dt + b(Y_t;\theta)dX_t, \quad Y_0 = y_0 \tag{4.14}$$

where $f(Y_t;\theta) : \mathbb{R}^m \to \mathbb{R}^m$ is a Lipschitz function, $b(Y_t;\theta) : \mathbb{R}^m \to \mathbb{R}^m$ is a γ-Lipschitz function with $\gamma > p$ as defined in Definition 5. Let $\{X_t, t \in [0,T]\}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathcal{F}_{(t>0)}, P)$ which can be expressed as the limit of its piecewise linear interpolation in p-variation metric almost surely, and by the Universal Limit Theorem, the solution $\{Y_t, t \in [0,T]\}$ can be lifted to a geometric p-rough path.

Assume that the observations of $\{Y_t\}$ are made on the homogeneous time point set

$$\mathcal{D}_n = \{t_0, t_1, ..., t_n\} = \{0, \delta, ..., n\delta\},$$

with n the number of observations and $\delta := \frac{T}{n}$, the time between observations being fixed. The observations are denoted by $Y^{\mathcal{D}_n}$, and the aim is to construct approximate MLEs for θ from the observations $Y^{\mathcal{D}_n}$. The likelihood function of (4.14) depends on the continuous trajectory of the path, and thus the exact likelihood function can not be obtained by only using observations at discrete time points. Thus, we introduce a method which is easy to apply. The method is to construct an approximate likelihood function by building an approximate model. The likelihood function of the approximate model is taken as the approximate likelihood function. In order to construct the approximate model, we first construct a piecewise linear path, $\{X(n)_t\}$ as follows, for $t \in [t_i, t_{i+1})$,

$$X(n)_t = X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{\delta}(t - t_i), \tag{4.15}$$

and an ordinary differential equation can be constructed for, as follows

$$dY(n)_t = f(Y(n)_t; \theta)dt + b(Y(n)_t; \theta)dX(n)_t, \tag{4.16}$$

and the initial condition is $Y(n)_0 = y_0$. The solution on $\mathcal{D}_n$ is denoted by

$$Y(n)^{\mathcal{D}_n} = \{Y(n)_{t_i}, t_i \in \mathcal{D}_n\}.$$

We take the above model as the approximation of model (4.14), and use the likelihood function of the above model as the approximate likelihood of (4.14). The original model is referred to as the limiting model and the approximate model is referred to as the discrete model.

We provide some justification on the method. Let $L_{Y(n)}(\theta|.)$ be the likelihood function of $Y(n)^{\mathcal{D}_n}$ and $\hat{\theta}_T(.)$ the corresponding MLEs. $\hat{\theta}_T(Y^{\mathcal{D}_n})$ is calculated by plugging in the observations from the limiting model and $\hat{\theta}_T(Y(n)^{\mathcal{D}_n})$ is calculated by plugging in realisations of the discrete model which is driven by the same driving force as the limiting model. We have the following

$$|\theta - \hat{\theta}_T(Y^{\mathcal{D}_n})| \leq |\hat{\theta}_T(Y(n)^{\mathcal{D}_n}) - \theta| + |\hat{\theta}_T(Y^{\mathcal{D}_n}) - \hat{\theta}_T(Y(n)^{\mathcal{D}_n})|. \tag{4.17}$$

The first part of the upper bound is controlled if the MLE of the approximate process is consistent. The second part of the upper bound is due to the data-model mismatch error, and one of the sufficient condition for the error to be controlled is that

$$|L_{Y(n)}(\theta|Y^{\mathcal{D}_n}) - L_{Y(n)}(\theta|Y(n)^{\mathcal{D}_n})| < Cd_p(\mathbf{Y}, \mathbf{Y}(n))_{[0,T]},$$

where $\mathbf{Y}$ and $\mathbf{Y}(n)$ the geometric rough path lifted by $\{Y_t\}$ and $\{Y(n)_t\}$, and C is a constant independent of the path, and the parameters, but might depend on δ. From above, we would like to study the continuity property of the log-likelihood function in the next section, and we focus ourselves on the processes driven by fractional Brownian motion.

4.4 Asymptotic behaviour of the error of the log-likelihood

In this section, we study the continuity of the log-likelihood function of the fractional process.

4.4.1 Formulation of the problem

In this section, we construct the exact log-likelihood function for the discrete model (4.16), and then discuss the asymptotic behaviour of the log-likelihood function with respect to the observations, that is, as the time gap between observations goes to zero, whether the log-likelihood function will converge or not. We also refer to as the convergence property of the log-likelihood function. We constrain ourselves to the one-dimensional process. Let $\{X(n)_t\}$ be the driving force of the discrete model. From (4.15), it can be parametrised by the random vector

$$\Delta X(n)^{\mathcal{D}_n} = [\Delta X(n)_{t_1}, \Delta X(n)_{t_2}, ..., \Delta X(n)_{t_n}]^*,$$

where $\Delta X(n)_{t_i} = X(n)_{t_i} - X(n)_{t_{i-1}}$, and $*$ is the conjugate operator.

Remark A remark is made for the use of conjugate operator. We use this operator to denote the transpose of the real matrix instead of T, since we want to avoid confusion on notations, where T refers to the time throughout the thesis.
Given a set of observations $Y(n)^{\mathcal{D}_n} = \{y(n)_0, y(n)_{t_1}, ..., y(n)_{t_n}\}$ generated by (4.16), we need to express $\Delta X(n)^{\mathcal{D}_n}$ by the observations $Y(n)^{\mathcal{D}_n}$ in order to construct the log-likelihood function. We introduce the procedure which is given in [52]. For

$t \in (t_{i-1}, t_i]$, we have

$$dY(n)_t = f(Y(n)_t;\theta) + b(Y(n)_t;\theta)\frac{\Delta X(n)_{t_i}}{\delta}dt, \qquad Y(n)_{t_{i-1}} = y(n)_{t_{i-1}}$$

and we denote the above solution map by $Y(n)_t = F_{t-t_{i-1}}(y(n)_{t_{i-1}}, \Delta X_{t_i};\theta)$. We consider such parameters θ that the solution of the above differential equation exists. In order for $\{Y(n)_t\}$ to go through the observations, ΔX_{t_i} should satisfy the following equation,

$$F_{t_i-t_{i-1}}(y(n)_{t_{i-1}}, \Delta X_{t_i};\theta) = y(n)_{t_i}.$$

Let $\Delta X_{t_i}(y(n)_{t_{i-1}}, y(n)_{t_i};\theta)$ be the solution for the above equation, and we define a map $I^{-1}_{y_0,\theta}(Y(n)^{\mathcal{D}_n}) = \{\Delta X_{t_i}(y(n)_{t_{i-1}}, y(n)_{t_i};\theta), t_i \in \mathcal{D}_n\}$. Thus, the Jacobian matrix $\nabla I^{-1}_{\theta,y_0}$ is an n by n lower triangular matrix, with the diagonal being $\frac{\partial \Delta X_{t_i}(y_{t_{i-1}}, y_{t_i};\theta)}{\partial y_{t_i}}$.
Since the driving force is the piecewise-linear interpolation of a fractional Brownian motion which is a Gaussian process, the likelihood function is

$$\begin{aligned} L(\theta|Y(n)^{\mathcal{D}_n}) &= f(I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n}))|\nabla I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})| \\ &= \frac{1}{\sqrt{(2\pi)^n|\Sigma_n|}} e^{-\frac{1}{2}I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})^*\Sigma_n^{-1}I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})}|\nabla I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})|, \end{aligned}$$

where Σ_n is the covariance matrix of the increments of the fractional Brownian motion, and f is the probability density for fractional noise. The log-likelihood function is

$$l_T(\theta|Y(n)^{\mathcal{D}_n}) = log(\frac{1}{\sqrt{(2\pi)^n|\Sigma_n|}}) - \frac{1}{2}I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})^*\Sigma_n^{-1}I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n}) + log|\nabla I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})|. \tag{4.18}$$

Next, we discuss the convergence of the log-likelihood function. Recall that $Y^{\mathcal{D}_n}$ is the observations from the limiting model (4.14) on $\mathcal{D}_n$, and the difference between

$l_T(\theta|Y^{\mathcal{D}_n})$ and $l_T(\theta|Y(n)^{\mathcal{D}_n})$ is

$$\begin{aligned}
&|l_T(\theta|Y(n)^{\mathcal{D}_n}) - l_T(\theta|Y^{\mathcal{D}_n})| = |\frac{1}{2} I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})^* \Sigma_n^{-1} I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n}) - \frac{1}{2} I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})^* \Sigma_n^{-1} I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n}) \\
&+ log|\nabla I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})| - log|\nabla I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})|| \\
&\leq \frac{1}{2}|I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})^* \Sigma_n^{-1} I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n}) - I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})^* \Sigma_n^{-1} I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})| \\
&+ |log|\nabla I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})| - log|\nabla I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})||.
\end{aligned} \tag{4.19}$$

We want to make some remark on $I^{-1}_{\theta,y_0}(.)$. $I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})$ yields a sequence of increments $\Delta X^{\mathcal{D}_n}$, and we define a process $\{X(n)_t\}$ as follows, for $t \in [t_i, t_{i+1})$,

$$X(n)_t = X(n)_{t_i} + \frac{\Delta X^{\mathcal{D}_n}_{i+1}}{\delta}(t - t_i). \tag{4.20}$$

$I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})$ also yields the the sequence of increments $\Delta \tilde{X}^{\mathcal{D}_n}$, and we construct path $\{\tilde{X}(n)_t\}$ as follows,

$$\tilde{X}(n)_t = \tilde{X}(n)_{t_i} + \frac{\Delta \tilde{X}^{\mathcal{D}_n}_{i+1}}{\delta}(t - t_i). \tag{4.21}$$

As we can see that the distance is controlled by two parts, the quadratic form and the Jacobian part. We discuss them separately.

Control of the Jacobian The convergence of the Jacobian is discussed in [52] (Lemma 7.2), and we cite the result as follows

Theorem 4.4.1. *[52] For the model (4.14), assume that $f(.;\theta)$ and $b(.;\theta)$ are both $\gamma+1$-Lipschitz continuous uniformly in θ, for some $\gamma > p$. And for $\mathcal{M}_b > 0$,*

$$\inf_{y,\theta} \|b(y;\theta)\| = \frac{1}{\mathcal{M}_b} \tag{4.22}$$

the Jacobian is controlled as

$$|log|\nabla I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})| - log|\nabla I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})|| \leq Cw(d_p(X(n), \tilde{X}(n))), \tag{4.23}$$

for some $C \in \mathbb{R}^+$, and modulus of continuity function w.

Recall that the γ-Lipschitz is defined in Definition 5, and p is the level of roughness of the paths.

Control of the quadratic form With the above assumption of the Jacobian, the continuity of the log-likelihood function is equivalent to the continuity of the quadratic form

$$\begin{aligned}&|I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n})^*\Sigma_n^{-1}I^{-1}_{\theta,y_0}(Y(n)^{\mathcal{D}_n}) - I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})^*\Sigma_n^{-1}I^{-1}_{\theta,y_0}(Y^{\mathcal{D}_n})|\\ &= |\Delta X^{\mathcal{D}_n *}\Sigma_n^{-1}\Delta X^{\mathcal{D}_n} - \Delta\tilde{X}^{\mathcal{D}_n *}\Sigma_n^{-1}\Delta\tilde{X}^{\mathcal{D}_n}|.\end{aligned} \tag{4.24}$$

Due to the symmetry of Σ_n^{-1}, for $\frac{1}{h} \leq 2$ and $p > \frac{1}{h}$, we have for (4.24) as follows,

$$\begin{aligned}&|\Delta X^{\mathcal{D}_n *}\Sigma_n^{-1}\Delta X^{\mathcal{D}_n} - \Delta\tilde{X}^{\mathcal{D}_n *}\Sigma_n^{-1}\Delta\tilde{X}^{\mathcal{D}_n}|\\ &= |(\Delta X^{\mathcal{D}_n} + \Delta\tilde{X}^{\mathcal{D}_n})\Sigma_n^{-1}(\Delta X^{\mathcal{D}_n} - \Delta\tilde{X}^{\mathcal{D}_n})|\\ &\leq \|\Sigma_n^{-1}\|_2\|\Delta X^{\mathcal{D}_n} + \Delta\tilde{X}^{\mathcal{D}_n}\|_2\|\Delta X^{\mathcal{D}_n} - \Delta\tilde{X}^{\mathcal{D}_n}\|_2\\ &\leq \|\Sigma_n^{-1}\|_2\|\Delta X^{\mathcal{D}_n} + \Delta\tilde{X}^{\mathcal{D}_n}\|_2 d_p(X(n), \tilde{X}(n))_{[0,T]},\end{aligned} \tag{4.25}$$

where $\|.\|_2$ for a vector is the l^2 norm, and $\|.\|_2$ for a matrix is defined as $\sup_x \frac{\|Ax\|_2}{\|x\|_2}$ with x a non-zero vector.

Conclusion for the convergence of the log-likelihood function From (4.23) and (4.25), we need $d_p(X(n), \tilde{X}(n))_{[0,T]} \to 0$. A.Papavasiliou and K.Taylor prove the convergence result in ([52]), as follows

$$\lim_{n\to\infty} \sup_H d_p(X(n), \tilde{X}(n))_{[0,T]} = 0,$$

with Assumption 4.22. The driving force depends on the parameters θ, and H is the range of the parameters.

Thus, from (4.25), the Jacobian is convergent with respect to the observations as $n \to \infty$ and T is fixed.

For the quadratic part, from (4.25), we need to investigate the 2-norm of Σ_n^{-1} at least for the case when $p < 2$. Since Σ_n^{-1} is symmetric, its 2-norm equals to the maximum eigenvalue [44]. It is known that Σ_n is a positive definite matrix, and thus Σ_n^{-1} has finite 2-norm[44]. However, the eigenvalues of Σ_n can approach to zero as $n \to \infty$, which means that the eigenvalues of Σ_n^{-1} can diverge as $n \to \infty$. Therefore it is necessary to investigate the behaviour of the 2-norm of the matrix Σ_n^{-1} as $n \to \infty$.

First, we make an observation that the covariance matrix is a Toeplitz matrix. Recall from (3.1) that Σ_n, the covariance matrix of the increment of fractional

Brownian motion on $\mathcal{D}_n$ is as follows,

$$\Sigma_n = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-2} & \rho_{n-3} & \cdots & \cdots & \rho_1 \\ \rho_{n-1} & \rho_{n-2} & \cdots & \rho_1 & \rho_0 \end{bmatrix}.$$

Toeplitz matrix is defined as the matrix such that the entities are of the form $T_{i,j} = T_{i-j}$ [29]. As we can see, $\Sigma_n(i,j) = \rho_{i-j}$, and thus it is a Toeplitz matrix. The covariance matrix can be considered as an operator describing the covariance structure of such fractional processes $\int_0^1 f_n(t)dB_t^h$ where f_n is a step function of n-th level. In the next section, we provide a brief literature review on the research of asymptotic behaviour of eigenvalues of the Toeplitz matrix and covariance operator of fractional Brownian noise.

4.4.2 Preliminaries on Toeplitz form

As discussed from the previous section, the continuity of the log-likelihood function depends on the lower bound of the eigenvalues of the covariance matrix which is a Toeplitz matrix. Hence, we provide some preliminaries and some literature review on the asymptotic behaviour of the eigenvalues of the Toeplitz matrix.
We introduce Szegö's theory on Toeplitz form in this section and the material is from the monograph on Toeplitz form (Chapter 5, p62-66) [29].
Let $A(n)$ be an n-dimensional Toeplitz matrix with $A(n)_{k,l} = a_{k-l}$, and $f(x)$ be a real valued Lebesgue measurable function such that

$$a_{k,l} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i(k-l)x} f(x)dx, \quad k-l = 0, \pm 1, \pm 2, ...$$

Then $T_n(f) := A(n)$ is the finite Toeplitz form generated by $f(x)$ and $f(x)$ is called the symbol of the finite Toeplitz form. We only consider the case where $A(n)$ is a real valued matrix, and thus $a_{k-l} = a_{l-k}$, and $f(x)$ is an even function.
Let m be the essential lower bound of $f(x)$ which is the largest lower bound of $f(x)$ on $[-\pi, \pi] \setminus \eta$ where η is a set of Lebesgue measure 0, and M be the the essential upper bound of $f(x)$ which is the smallest upper bound of $f(x)$ on $[-\pi, \pi] \setminus \eta$. We

denote the eigenvalues of $T_n(f)$ as follows in a non-descending order, that is

$$\lambda_1^{(n)} \leq \lambda_2^{(n)} \leq ... \leq \lambda_n^{(n)},$$

and G.Szegö [29] proved that

$$m \leq \lambda_v^{(n)} \leq M, \quad 1 \leq v \leq n \tag{4.26}$$

We are interested in the asymptotic behaviour of the eigenvalues of the Toeplitz matrix $A(n)$, that is the eigenvalues of Toeplitz form defined as

$$T(f) = \lim_{n\to\infty} T_n(f).$$

G.Szegö's theorem on the eigenvalues of the Toeplitz form $T(f)$ is as follows,

Theorem 4.4.2. *[29] Let $f(x)$ be a real valued function which is Lebesgue measurable, and $\lambda_i^{(n)}$ be the eigenvalues of $T_n(f)$. We denote by m and M the 'essential' lower bound and upper bound of $f(x)$ respectively. If $F(\lambda)$ is any continuous function defined in the interval $m \leq \lambda \leq M$, we have*

$$\lim_{n\to\infty} \frac{F(\lambda_1^{(n)}) + F(\lambda_2^{(n)}) + ... + F(\lambda_n^{(n)})}{n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(f(x))dx.$$

It is worth pointing out that the theorem is still valid even if m and M is infinite. From the above theorem, Segö [29] stated the following result,

$$\lim_{n\to\infty} \lambda_1^{(n)} = m, \quad \lim_{n\to\infty} \lambda_n^{(n)} = M. \tag{4.27}$$

By Szegö's theorem, the bounds of the 2-norm (eigenvalues) of Σ_n are within the range of the symbol $f(x)$. Thus, the proof of the boundedness for the 2 norm of the covariance matrix consists two steps. First, we prove that there exists a measurable function on $[-\pi, \pi]$ such that its Fourier coefficients are $\{\rho_i, i \in \mathbb{N}\}$. Then, we need to find the range of this function. We divide the discussion into two cases which are $h < \frac{1}{2}$ and $h > \frac{1}{2}$.

The study of the asymptotic behaviour of eigenvalues of the Toeplitz matrix is based on Szegö's theory on Toeplitz form. An important concept of the theory is the symbol of the Toeplitz matrix. Let A_n be an n-dimensional Toeplitz matrix with $A_n(i,j) = a_{i,j} = a_{|i-j|}$. Let $f(x)$ be a Lebesgue measurable real valued function on

$[-\pi, \pi]$ such that

$$a_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikx} f(x)dx$$

are its Fourier coefficients. The function $f(x)$ is refereed to as the symbol of the matrix A_n[29].

The studies focus on the application of the theory to the Teoplitz forms generated by specific symbols. For instance, in [18], Dai et al discussed the asymptotic behaviour of eigenvalues and eigenvectors of Fisher–Hartwig matrix, and in [10], Bürger gave the calculation of the inverse and determinate of the linear growth real Toeplitz matrix where the elements is of the the form $A_{i,j} = a + b|i - j|$.

Another topic is the asymptotic behaviour of the extreme eigenvalues. In [60], Serra proved that the speed of the convergence of the smallest eigenvalue to the minimum f_m of the symbol f only depends on the speed of f converging to f_m. That is $|f - f_m| \sim O(|\lambda_0^n - f_m|)$, where λ_0^n is the smallest eigenvalue.

We would like to introduce the works on the eigenvalue problem of the covariance operator of the fractional Brownian noise [14]. The operator is an operator such that it defines the relation of the covariance of the stochastic integral of deterministic functions [14]

$$\mathbb{E}(\int_0^1 f(s)dB_s^h \int_0^1 g(s)dB_s^h) = \int_0^1 f(s)(Kg)(s)ds,$$

and it is explicitly given by [14]

$$(Kf)(t) = \frac{d}{dt}\int_{s=0}^{1} h(t-s)^{2h-1} sign(t-s) f(s)ds.$$

Then the eigenvalues of the operator K are approximated as follows [14],

$$\lambda_n = ((n - \frac{1}{2})\pi + \frac{1-2h}{4}\pi + O(n^{-1}))^{1-2h},$$

and as we can see, the eigenvalues explode as $h < \frac{1}{2}$, since the covariance operator for the fractional Brownian noise is not compact, and for $h > \frac{1}{2}$, $\lambda_n \to 0$ since K is compact [14].

Let Σ_n be the covariance matrix of the increment of fractional Brownian motion, and the eigenvalues of the covariance matrix of the increment of fractional Brownian

motion are defined by $\Sigma_n x = \lambda x$. It can be considered as the analogy of the eigenvalues of the covariance operator. However, the behaviour of the eigenvalues of Σ_n are different to those of the covariance operator of fractional Brownian noise. As we will see in the following sections, the eigenvalues of Σ_n converge to zero as $n \to \infty$ for $h < \frac{1}{2}$, and diverge as $\frac{1}{2} < h < 1$. Heuristically, the reason for the divergence of the eigenvalues of K with $h < \frac{1}{2}$ is that $K(s,s)$ is divergent and the divergent speed is faster as h is smaller. However, when we fix δ, Σ_n does not diverge along the diagonal and $\Sigma_n(i,j)$ is smaller as h is smaller. Thus, it is reasonable that the eigenvalues of Σ_n are smaller when h is smaller. Therefore, the eigenvalues of Σ_n and K behave in the opposite direction. In the next section, we study the eigenvalues of Σ_n.

4.4.3 Uniform Boundedness of Σ_n^{-1}

Let $\lambda_{max}(\Sigma_n^{-1})$ denote the maximum eigenvalue of Σ_n^{-1}, and since Σ_n^{-1} is symmetric, $\|\Sigma_n^{-1}\|_2 = \lambda_{max}(\Sigma_n^{-1})$. In addition, because [44] $\lambda_{max}(\Sigma_n^{-1}) = \frac{1}{\lambda_{min}(\Sigma_n)}$, the goal is to study the lower bound of the minimum of the eigenvalues of Σ_n as $n \to \infty$. The main theory used to study Σ_n is Szegö's theory on Toeplitz form, and we provide some preliminaries on the theory.

4.4.3.1 Case 0.25<h<0.5

We make an remark on the range of the Hurst index. For the rough differential equation to admit solutions, we require that the paths are geometric p-rough paths. By [17], the fractional Brownian motion is geometric rough path only when $h > 0.25$, and thus we focus ourselves on the range where $h > 0.25$ throughout the thesis.

We normalise the covariance matrix by considering the process on integer time, that is

$$\{t_0, t_1, ..., t_n\} = \{0, 1, ..., n\},$$

since δ, the time between consecutive observations is fixed. In this section, we consider the case where $h < 0.5$. By (3.2), for $h < 0.5$,

$$\sum_{k=0}^{\infty} |\rho_k| < \infty,$$

and thus there exits a measurable function $f(x)$ on $[-\pi, \pi]$

$$f(x) = \lim_{n\to\infty} S_n(x) := \sum_{-\infty}^{\infty} \rho_k cos(kx)$$

such that its Fourier coefficients are $\{\rho_k, k \in \mathbb{Z}\}$ by the Dominated Convergence Theorem (A.1.5). Then since ρ_k is negative, and $|coskx| \leq 1$, $\rho_k coskx \geq \rho_k$ for every k, and the minimum value of $f(x)$ on $[-\pi, \pi]$ is $f(0)$.
Consider the following,

$$\begin{aligned}
&S_n(0) = \rho_0 + 2\sum_{k=1}^{n} \rho_k = 1 + 2\sum_{k=1}^{n} \mathbb{E}(\Delta X_{t_0}, \Delta X_{t_i}) = 1 + 2\mathbb{E}(\Delta X_{t_0}, (X_{t_n} - X_{t_1})) \\
&= 1 + 2(\mathbb{E}(X_{t_1} X_{t_n}) + \mathbb{E}(X_{t_0} X_{t_1}) - \mathbb{E}(X_{t_1}^2) - \mathbb{E}(X_{t_0} X_{t_n})) \\
&= 1 + (t_1^{2h} + t_n^{2h} - (t_n - t_1)^{2h} + t_0^{2h} + t_1^{2h} - (t_1 - t_0)^{2h} - 2t_1^{2h} - t_0^{2h} - t_n^{2h} + \frac{1}{2}(t_n - t_0)^{2h}) \\
&= (t_n - t_0)^{2h} - (t_n - t_1)^{2h} = (n^{2h} - (n-1)^{2h}) \\
&= ((n-1)^{2h} + 2h(n-1)^{2h-1} + O((n-1)^{2h-2})) - (n-1)^{2h} \sim O(n^{2h-1})
\end{aligned} \tag{4.28}$$

The last three steps are obtained by applying Taylor expansion to n^{2h} at $n-1$. Therefore as n goes to infinity, the minimum eigenvalue of the covariance matrix Σ_n converges to $f(0) = \lim_{n\to\infty} S_n(0) = 0$ by (4.27). Hence Σ_n^{-1} does not have a uniformly bounded 2 norm with respect to n. We make a discussion on the speed of the convergence. Let $f_n(x) := S_n(x) = \sum_{-(n-1)}^{n-1} \rho_k cos(kx)$, and thus the Fourier coefficients of f_n are $\Sigma_n(i-j)$, by the fact

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-kix} f_n(x)dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-kix} \sum_{k=-(n-1)}^{n-1} \rho_k e^{ikx} dx = \rho_k.$$

From (4.26)[29] and (4.28), we have

$$\lambda_1^{(n)} \geq \min_{[-\pi,\pi]} f_n(x) = f_n(0) \sim O(n^{2h-1})$$

Hence, a lower bound of the convergence rate of the minimum of the eigenvalues of Σ_n is $O(n^{2h-1})$, and thus $\|\Sigma_n^{-1}\|_2$ explodes at a speed of $O(n^{1-2h})$.

4.4.3.2 Case h>0.5

In this section, we consider the case where $h > 0.5$, and we first introduce a theorem. A sequence $(a_k)_{k\geq 0}$ is convex if $\Delta^2 a_k \geq 0$ for all $k \geq 0$ where $\Delta^2 a_k = \Delta a_k - \Delta a_{k+1}$ and $\Delta a_k = a_k - a_{k+1}$. It is strictly convex if the inequality is strict. First, we introduce a theorem regarding to the convex sequence as follows [68](Chapter 5, Theorem 1.5, p183),

Theorem 4.4.3. *Let $(a_k)_{k\geq 0}$ be a convex sequence with $a_k \to 0$, and then the series*

$$S_n(x) = a_0 + 2\sum_{i=1}^{n} a_i cos(ix),$$

converges in a pointwise sense to a non-negative and integrable function $f(x) : [-\pi, \pi] \setminus \{0\} \to \mathbb{R}$ whose Fourier coefficients are a_k.

First, we prove that

Observation 4.4.1. *The sequence $\{\rho_i, i \in \mathbb{N}\}$ is strictly convex when $h > \frac{1}{2}$.*

Proof. The entries of Σ_n are as follows,

$$\rho_i = \begin{cases} 1 & i = 0 \\ 0.5((i+1)^{2h} + (i-1)^{2h} - 2i^{2h}) & i \geq 1 \end{cases}$$

First, we consider the sequence with the index $i \geq 1$. The convexity of the sequence $\{\rho_i, i \in \mathbb{N}^+\}$ is equivalent to the convexity of function $a(t) = 0.5((t+1)^{2h} + (t-1)^{2h} - 2t^{2h})$ where $t \in \mathbb{N}^+$. The second derivative of $a(t)$ is

$$a(t)'' = h(2h-1)\{(t+1)^{2h-2} + (t-1)^{2h-2} - 2t^{2h-2}\}.$$

Since $g(x) = x^{2h-2}$ for $x \geq 1$ is a strictly convex function, $(x+1)^{2h-2} + (x-1)^{2h-2} - 2x^{2h-2} > 0$. Thus, $a(t)'' > 0$ for $t \geq 1$. Therefore $a(t)$ is a strictly convex function for $t \geq 1$, and so is $\{\rho_i\}$ for $i \in \mathbb{N}^+$.
For the convexity of the sequence at $i = 0$, we have $\Delta^2\rho_0 = \rho_0 + \rho_2 - 2\rho_1 = 0.5 * 3^{2h} + 3.5 - 2 * 2^{2h}$ which is a function of h, and we denote it by $q(h)$. Taking the derivative of $q(h)$, we have $q'(h) = h2^{h-1}(1.5^{2h-1} - 4)$, and thus $q(h)$ is decreasing for $h > \frac{1}{2}$. Because $q(1) = 0$, and $q(h)$ is continuous, $q(h)$ is positive on $h \in (\frac{1}{2}, 1)$. Thus $\{\rho_i\}$ is strictly positive for $i \in \mathbb{N}$. □

By Theorem 4.4.3, there exists a non-negative function on $[-\pi, \pi] \setminus \{0\}$ defined as $f(x) = \lim_{n\to\infty}(\rho_0 + 2\sum_{i=1}^{n} \rho_i cos(ix))$, and we proceed to prove that $f(x)$ is positive on $x \in [-\pi, \pi]$. First, from [68], we should use summation by parts to rearrange the sequence. The summation by parts is as follows,

Theorem 4.4.4. *Given two sequences* $\{a_n\}$ *and* $\{b_n\}$ *for* $n = 0, 1, 2, ...,$ *and let* $A_n = \sum_{k=0}^{n} a_k$. *Then, we have*

$$\sum_{n=0}^{N} a_n b_n = \sum_{n=0}^{N-1} A_n(b_n - b_{n+1}) + A_N b_N.$$

Proof. Notice that,

$$A_{N-2}(b_{N-2} - b_{N-1}) + A_{N-1}(b_{N-1} - b_N) = A_{N-2}b_{N-2} + (A_{N-1} - A_{N-2})b_{N-1} - A_{N-1}b_N$$
$$= A_{N-2}b_{N-2} + a_{N-1}b_{N-1} - A_{N-1}b_N,$$

and for $n \leq N - 3$, we have

$$A_n(b_n - b_{n+1}) + A_{n+1}b_{n+1} = A_n b_n + (A_{n+1} - A_n)b_{n+1}$$
$$= A_n b_n + a_{n+1}b_{n+1}.$$

Thus, $\sum_{n=0}^{N-1} A_n(b_n - b_{n+1}) = \sum_{n=0}^{N-1} a_n b_n - A_{N-1}b_N$, and adding both sides by $A_N b_N$ completes the proof. □

Theorem 4.4.5. *The function* $f(x) = \lim_{n\to\infty}(\rho_0 + 2\sum_{i=1}^{n} \rho_i cos(ix))$ *is positive for* $x \in [-\pi, \pi]$.

The proof is mainly from the proof of Theorem4.4.3[68](Chapter 5, Theorem 1.5, p183).

Proof. Let $S_n(x) = \rho_0 + 2\sum_{i=1}^{n} \rho_i cos(ix)$, and by applying summation by parts twice, we can transform the series $S_n(x)$ in the following way using Theorem 4.4.4 [68],

$$S_n(x) = \sum_{i=-n}^{n} \rho_i cos(ix) = \sum_{i=0}^{n-1} D_i(x)\Delta\rho_i + D_n(x)\rho_n$$
$$= \sum_{i=0}^{n-2}(i+1)K_i(x)\Delta^2\rho_i + nK_{n-1}(x)\Delta\rho_{n-1} + D_n(x)\rho_n,$$

where $D_n(x)$ is the Dirichlet's kernel [68],

$$D_n(x) = \sum_{k=-n}^{n} e^{ikx} = \frac{sin(x(n+0.5))}{sin(0.5x)}.$$

and $K_n(x)$ is the Fejér's kernel [68],

$$K_n(x) = \frac{1}{1+n}\sum_{k=0}^{n} D_k(x) = \frac{1}{1+n}\frac{sin^2(0.5x(n+1))}{sin^2(0.5x)}.$$

We consider the last two terms in the above equation. For $x \in [-\pi, \pi] \setminus \{0\}$,

$$|nK_{n-1}(x)\Delta\rho_{n-1}| = \frac{sin^2(0.5x(n+1))}{sin^2(0.5x)}\Delta\rho_{n-1} \leq \frac{1}{sin^2(0.5x)}\Delta\rho_{n-1}.$$

Since

$$\lim_{n\to\infty} \Delta\rho_{n-1} = \lim_{n\to\infty} \rho_{n-1} - \lim_{n\to\infty} \rho_n = 0,$$

and $\frac{1}{sin^2(0.5x)} < \infty$ for $x \in [-\pi, \pi] \setminus \{0\}$, we have $nK_{n-1}(x)\Delta\rho_{n-1} \to 0$. For the other term, we have

$$|D_n(x)\rho_n| = |\frac{sin(x(n+0.5))}{sin(0.5x)}\rho_n| \leq \frac{1}{|sin(0.5x)|}\rho_n.$$

Since $\frac{1}{sin(0.5x)} < \infty$ for $x \in [-\pi, \pi] \setminus \{0\}$, and by (3.2), $\rho_n \to 0$, $D_n(x)\rho_n \to 0$. Therefore,

$$f(x) := \lim_{n\to\infty} \sum_{v=0}^{n} (v+1)K_v(x)\Delta^2\rho_v.$$

For any $x \in [-\pi, \pi]\backslash 0$, there exists a v such that, $K_v(x) > 0$, and thus $K_v(x)\Delta^2\rho_v > 0$. Since other terms are non negative, $f(x)$ is positive. Then, we consider $f(x)$ at $x = 0$, which is

$$f(0) := \lim_{n\to\infty} (\rho_0 + 2\sum_{i=1}^{n} \rho_i) = \infty,$$

since by (3.2), $\rho_n \sim O(n^{2h-2})$. Thus, $f(x)$ is positive on $[-\pi, \pi]$. □

By Szegö's Theorem, $\lim_{n\to\infty} \lambda_{min}(\Sigma_n) > 0$, and thus $\lim_{n\to\infty} \lambda_{max}(\Sigma_n^{-1}) = \|\Sigma_n^{-1}\|_2 < \infty$. We make a summery on the main results in this section.

Theorem 4.4.6. *Let Σ_n be the covariance matrix of a sequence of increments of fractional Brownian motion $\Delta X = [\Delta X_{t_1}, ..., \Delta X_{t_n}]^*$ where $t_{i+1} - t_i = \delta$ is fixed.*

Then $\lim_{n\to\infty} \|\Sigma_n^{-1}\|_2 \sim O(n^{1-2h})$ *for* $h < 0.5$, *and* $\lim_{n\to\infty} \|\Sigma_n^{-1}\|_2 < \infty$ *for* $h > 0.5$.

4.4.4 Results on the error of the log-likelihood

Since $\|\Sigma_n^{-1}\|_2$ is uniformly bounded when $h > \frac{1}{2}$, the log-likelihood function is continuous for fractional processes in that case. We make an theorem about the result.

Theorem 4.4.7. *Consider the log-likelihood function of the one-dimensional model (4.16) where the driving force is the linear interpolation of an h-fractional Brownian motion. For* $h > \frac{1}{2}$, *given assumption 4.22, the error between two log-likelihood functions is controlled by the p-variation metric between observations, where* $p > \frac{1}{h}$ *as* $n \to \infty$ *and* T *is fixed.*

Proof. From equation (4.19), the difference between likelihood function is

$$\begin{aligned}
&|l_T(\theta|Y(n)^{\mathcal{D}_n}) - l_T(\theta|Y^{\mathcal{D}_n})| \\
&\leq \frac{1}{2}|I_{\theta,y_0}^{-1}(Y(n)^{\mathcal{D}_n})^*\Sigma_n^{-1}I_{\theta,y_0}^{-1}(Y(n)^{\mathcal{D}_n}) - I_{\theta,y_0}^{-1}(Y^{\mathcal{D}_n})^*\Sigma_n^{-1}I_{\theta,y_0}^{-1}(Y^{\mathcal{D}_n})| \\
&+ |log|\nabla I_{\theta,y_0}^{-1}(Y(n)^{\mathcal{D}_n})| - log|\nabla I_{\theta,y_0}^{-1}(Y^{\mathcal{D}_n})||.
\end{aligned}$$

By Theorem 4.4.1, the Jacobian is controlled if given (4.22). For the quadratic part, from (4.25), we have, for $h > \frac{1}{2}$

$$\begin{aligned}
&|\Delta X^{\mathcal{D}_n *}\Sigma_n^{-1}\Delta X^{\mathcal{D}_n} - \Delta\tilde{X}^{\mathcal{D}_n *}\Sigma_n^{-1}\Delta\tilde{X}^{\mathcal{D}_n}| \\
&\leq \|\Sigma_n^{-1}\|_2\|\Delta X^{\mathcal{D}_n} + \Delta\tilde{X}^{\mathcal{D}_n}\|_2 d_p(X(n), \tilde{X}(n))_{[0,T]}.
\end{aligned}$$

From Theorem 4.4.6, $\|\Sigma_n^{-1}\|_2 < \infty$, and both $\{\tilde{X}(n)_t\}$ and $\{X(n)_t\}$ are paths of finite $\frac{1}{h}$-variation, thus finite 2-variation. Thus $\|\Delta X^{\mathcal{D}_n} + \Delta\tilde{X}^{\mathcal{D}_n}\|_2 < \infty$, and the quadratic part is controlled by the p-variation metric between $d_p(X(n), \tilde{X}(n))_{[0,T]}$ which is determined by $I_{\theta,y_0}^{-1}(Y(n)^{\mathcal{D}_n})$ and $I_{\theta,y_0}^{-1}(Y^{\mathcal{D}_n})$. By the continuity of the map $I_{\theta,y_0}^{-1}(.)$, the result is proved. □

4.5 Parameter estimation for the discretely observed Fractional O-U process

In this section, we apply the method to the one-dimensional fractional O-U process

$$dY_t = -\lambda Y_t dt + \sigma dX_t, \quad Y_0 = y_0, \tag{4.29}$$

where $\sigma, \lambda \in \mathbb{R}^+$, and $\{X_t, t \in (-\infty, T]\}$ is a fractional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_{(t>0)}, P)$. We adopt the invariant distribution assumption, and the initial condition is

$$y_0 = \int_{-\infty}^{0} e^{\lambda s} dX_s, \tag{4.30}$$

which exists almost surely under P [12] in pathwise sense. The differential equation has a strong unique solution for $h \in (0, 1]$ given by (see e.g [12]),

$$Y_t = Y_0 + \sigma \int_0^t e^{-(t-s)\lambda} dX_s, \tag{4.31}$$

and the integral is defined as pathwise Riemann-Stieljes integral.
Recall that δ is the fixed time between observations, and $T = n\delta$ is our observation duration. $\{Y_t\}$ is observed discretely at

$$\mathcal{D}_n := \{0, \delta, 2\delta, ..., n\delta\} = \{0, t_1, t_2, ..., t_{n-1}, t_n\} \tag{4.32}$$

and the observation of $\{Y_t\}$ on $\mathcal{D}_n$ is denoted by $Y^{\mathcal{D}_n}$.
Our aim is to construct approximate MLEs for λ and σ based on $Y^{\mathcal{D}_n}$, for which purpose, we introduce the approximate model. By the invariant distribution assumption, we define a time point set

$$\mathcal{D} := \{n\delta, (n-1)\delta, ..., 0, -\delta, -2\delta, ...\} = \{t_n, t_{n-1}, ..., 0, t_{-1}, t_{-2}, ...\},$$

and we construct $\{X(\delta)_t, t \in (-\infty, T]\}$ such that it is a piecewise linear interpolation of the control $\{X_t, t \in (-\infty, T]\}$ in (4.29),

$$X(\delta)_t = X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{\delta}(t - t_i), \tag{4.33}$$

for $t \in [t_i, t_{i+1})$, with $t_i, t_{i+1} \in \mathcal{D}$. $\{Y(\delta)_t, t \in [0,T]\}$ is constructed such that it satisfies the following differential equation,

$$dY(\delta)_t = -\lambda Y(\delta)_t dt + \sigma dX(\delta)_t, \tag{4.34}$$

and by the invariant distribution assumption, the initial condition is as follows

$$Y(\delta)_0 = \frac{\sigma(1-e^{-\lambda\delta})}{\lambda\delta} \sum_{k=0}^{\infty} e^{-k\delta\lambda}(X(\delta)_{t_{-k}} - X(\delta)_{t_{-k-1}}). \tag{4.35}$$

Thus, $\{Y(\delta)_t, t \in [0,T]\}$ is a piecewise smooth path. For convenience, we refer to model (4.29) as the limiting model and (4.34) as the discrete model. Recall from (4.34) that the discrete fractional O-U process is

$$dY(\delta)_t = -\lambda Y(\delta)_t dt + \sigma dX(\delta)_t, \quad Y(\delta)_0 = y(\delta)_0$$

where $\{X(\delta)_t\}$ is a piecewise linear random process on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and $y(\delta)_0$ satisfies (4.35). Suppose we have observations $\{Y(\delta)_{t_i}, t_i \in \mathcal{D}_n\}$, and δ, the time between observations is fixed. We would like to construct the MLEs for λ and σ using the observation $\{Y(\delta)_{t_i}, t_i \in \mathcal{D}_n\}$.
The section is organized as follows. First the likelihood function of discrete model is constructed and the MLEs for the drift coefficient and diffusion coefficient are derived. Then we prove the weak convergence of the MLEs for the discrete model.

4.5.1 Maximum Likelihood Estimator

Before the construction, we make further discussion on the path space of $\{Y(\delta)_t, t \in (-\infty, T]\}$.

Path space as an infinite dimensional vector space Notice that $\{Y(\delta)_t, t \in (-\infty, T]\}$ can be parametrised by a discrete process $\{Y(\delta)_t, t \in \mathcal{D}\}$, where

$$\mathcal{D} = \{t_n, t_{n-1}, ..., t_0, t_{-1}, ...\}, \tag{4.36}$$

and the discrete process $\{Y(\delta)_t, t \in \mathcal{D}\}$ can be considered as an infinite dimensional vector

$$Y(\delta)^{\mathcal{D}} = [Y(\delta)_{t_n}, Y(\delta)_{t_{n-1}}, ..., Y(\delta)_{t_0}, Y(\delta)_{t_{-1}} ...]^*.$$

Let $e_i \in \mathbb{R}^\infty$ with $i \in \mathbb{Z}^+$ be the basis

$$\begin{aligned} e_1 &= [1,0,0,0,0,\dots]^* \\ e_2 &= [0,1,0,0,0,\dots]^* \\ &\vdots \\ e_n &= [\underbrace{0,\dots,0}_{n-1},1,\dots]^* \\ &\vdots \end{aligned}$$

and for $v \in \mathbb{R}^\infty$, we have $v = \sum_{i=1}^{\infty} < v, e_i > e_i = \sum_{i=1}^{\infty} v_i e_i$, where

$$< x, y >= \sum_{i=1}^{\infty} x_i y_i,$$

for $x, y \in \mathbb{R}^\infty$. Thus, for the vector $Y(\delta)^{\mathcal{D}}$, we have

$$Y(\delta)^{\mathcal{D}} = \sum_{i=1}^{\infty} < Y(\delta)^{\mathcal{D}}, e_i > e_i = \sum_{i=1}^{\infty} Y(\delta)_{t_{n+1-i}} e_i.$$

We define a shift operator $S_{-k} : \mathbb{R}^\infty \to \mathbb{R}^\infty$, and we put a negative sign in front of k to indicate that the shift is backward along the time. For $k \in \mathbb{N}$, and $v \in \mathbb{R}^\infty$, we have

$$S_{-k}(v) = \sum_{i=1}^{\infty} < v, e_{i+k} > e_i = \sum_{i=1}^{\infty} v_{i+k} e_i.$$

Therefore, we have $S_{-k}(Y(\delta)^{\mathcal{D}}) = \sum_{i=1}^{\infty} < Y(\delta)^{\mathcal{D}}, e_{i+k} > e_i = \sum_{i=1}^{\infty} Y(\delta)_{t_{n+1-i-k}} e_i$. Let $b_i \in \mathbb{R}^n$ with $i \in \mathbb{Z}^+$ be the basis

$$\begin{aligned} b_1 &= [1,0,0,\dots,0]^* \\ b_2 &= [0,1,0,\dots,0]^* \\ &\vdots \\ b_n &= [\underbrace{0,\dots,0}_{n-1},1]^* \end{aligned}$$

We proceed to define the truncation of the shift operator, $S^n_{-k} : \mathbb{R}^\infty \to \mathbb{R}^n$, such that for $v \in \mathbb{R}^\infty$, $S^n_{-k}(v) = \sum_{i=1}^{n} < v, e_{i+k} > b_i = \sum_{i=1}^{n} v_{i+k} b_i$. Therefore, we have

$$S^n_{-k}(Y(\delta)^{\mathcal{D}}) = \sum_{i=1}^{n} < Y(\delta)^{\mathcal{D}}, e_{i+k} > b_i = \sum_{i=1}^{n} Y(\delta)_{t_{n+1-i-k}} b_i = [Y_{t_{n-k}}, Y_{t_{n-k-1}}, ..., Y_{t_{-k+1}}]^*.$$

We make a notation of the increment of $Y(\delta)^{\mathcal{D}}$ as follows,

$$\begin{aligned}\Delta Y(\delta)^{\mathcal{D}} = Y(\delta)^{\mathcal{D}} - S_{-1}(Y(\delta)^{\mathcal{D}}) &= \sum_{i=1}^{\infty} (< Y(\delta)^{\mathcal{D}}, e_i > - < Y(\delta)^{\mathcal{D}}, e_{i+1} >) e_i \\ &= \sum_{i=1}^{\infty} (Y(\delta)_{t_{n+1-i}} - Y(\delta)_{t_{n-i}}) e_i.\end{aligned}$$

By the same virtue, $\{X(\delta)_t, t \in (-\infty, T]\}$, the piecewise linear interpolation of $\{X_t, t \in (-\infty, T]\}$ on $\mathcal{D}$, can be parametrised by the vector

$$X(\delta)^{\mathcal{D}} = [X_{t_n}, X_{t_{n-1}}, ..., X_{t_0}, X_{t_{-1}} ...]^*,$$

and we define the increment of it as follows,

$$\begin{aligned}\Delta X(\delta)^{\mathcal{D}} = X(\delta)^{\mathcal{D}} - S_{-1}(X(\delta)^{\mathcal{D}}) &= \sum_{i=1}^{\infty} (< X(\delta), e_i > - < X(\delta), e_{i+1} >) e_i \\ &= \sum_{i=1}^{\infty} (X_{t_{n+1-i}} - X_{t_{n-i}}) e_i.\end{aligned}$$

Solve for $S^n_0(\Delta X(\delta)^{\mathcal{D}})$ In this section, we construct the likelihood function of the discrete model (4.34)

$$dY(\delta)_t = -\lambda Y(\delta)_t dt + \sigma dX(\delta)_t, \quad Y(\delta)_0 = y(\delta)_0,$$

First, we solve for the driving force $S^n_0(\Delta X(\delta)^{\mathcal{D}})$. For $t \in [t_i, t_{i+1})$, the partial differential equation can be transformed to the following ordinary differential equation,

$$\frac{dY(\delta)_t}{dt} = -\lambda Y(\delta)_t + \sigma \frac{X_{t_{i+1}} - X_{t_i}}{\delta},$$

and $Y(\delta)_t$ admits the following solution

$$Y(\delta)_t = \sigma \frac{X_{t_{i+1}} - X_{t_i}}{\delta} \int_{t_i}^{t} e^{-\lambda(t-s)} ds + Y(\delta)_{t_i} e^{\lambda(t_i - t)}. \tag{4.37}$$

Therefore, for $t = t_{i+1}$, we have

$$Y(\delta)_{t_{i+1}} = \frac{\sigma(1 - e^{-\lambda\delta})}{\lambda\delta} \left(X_{t_{i+1}} - X_{t_i}\right) + Y(\delta)_{t_i} e^{-\lambda\delta}, \tag{4.38}$$

and for vector $S_0^n(Y(\delta)^{\mathcal{D}})$, we have

$$S_0^n(Y(\delta)^{\mathcal{D}}) = \frac{\sigma(1 - e^{-\lambda\delta})}{\lambda\delta} S_0^n(\Delta X(\delta)^{\mathcal{D}}) + S_{-1}^n(Y(\delta)^{\mathcal{D}}) e^{-\lambda\delta}. \tag{4.39}$$

MLEs Recall from (4.18), the log-likelihood function of $Y(\delta)^{\mathcal{D}_n}$ is

$$l_T(\lambda, \sigma | Y(\delta)^{\mathcal{D}_n}) = log(\frac{1}{\sqrt{2\pi|\Sigma_n|}}) - \frac{1}{2} I_{\lambda,\sigma}^{-1}(Y(\delta)^{\mathcal{D}_n})^* \Sigma_n^{-1} I_{\lambda,\sigma}^{-1}(Y(\delta)^{\mathcal{D}_n}) + log|J_{I_{\lambda,\sigma}^{-1}}(Y(\delta)^{\mathcal{D}_n})|,$$

where Σ_n^{-1} is the inverse of the covariance matrix of the increment of the fractional Brownian motion 3.1. From (4.39), we have

$$I_{\lambda,\sigma}^{-1}(Y(\delta)^{\mathcal{D}_n}) = \frac{\lambda\delta}{\sigma(1 - e^{-\lambda\delta})} (S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}}) e^{-\lambda\delta}), \tag{4.40}$$

and from above, the Jacobian matrix is an n-dimensional lower triangular matrix with the diagonal elements, $\frac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}$, and thus the Jacobian is

$$|J_{I_{\lambda,\sigma}^{-1}}(Y(\delta)^{\mathcal{D}_n})| = \left(\frac{\lambda\delta}{\sigma(1 - e^{-\lambda\delta})}\right)^n. \tag{4.41}$$

Thus the log-likelihood function of $Y(\delta)^{\mathcal{D}_n}$ is

$$\begin{aligned}
&l_T(\lambda, \sigma | Y(\delta)^{\mathcal{D}_n}) \\
&= \frac{-n}{2} log(2\pi) - \frac{1}{2} log|\Sigma_n| - nlog\frac{\sigma(1 - e^{-\lambda\delta})}{\lambda\delta} \\
&- \frac{1}{2} \frac{\lambda^2\delta^2}{\sigma^2(1 - e^{-\lambda\delta})^2} (S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}}) e^{-\lambda\delta})^* \Sigma_n^{-1} (S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}}) e^{-\lambda\delta}).
\end{aligned}$$

The partial derivatives with respect to λ and σ are

$$\begin{cases} \frac{\partial l}{\partial \sigma} = -n\sigma^{-1} + \sigma^{-3}\frac{\lambda^2\delta^2(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})^*\Sigma_n^{-1}(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})}{(1-e^{-\lambda\delta})^2} \\ \frac{\partial l}{\partial \lambda} = \frac{1}{-2\sigma^2}\{\frac{2\lambda^2\delta^3(ae^{-2\lambda\delta} - be^{-\lambda\delta})}{(1-e^{-\lambda\delta})^2} \\ \quad +(S_0^n(Y(\delta)^{\mathcal{D}})) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})^*\Sigma_n^{-1}(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta}) \\ \quad \times(\frac{2\lambda\delta^2}{(1-e^{-\lambda\delta})^2} - \frac{2\lambda^2\delta^3e^{-\lambda\delta}}{(1-e^{-\lambda\delta})^3}) + n(\frac{1}{\lambda} - \frac{\delta e^{-\lambda\delta}}{1-e^{-\lambda\delta}}) \end{cases}$$

First, from $\frac{\partial l}{\partial \sigma} = 0$, we have

$$\hat{\sigma}_T^2(\delta) = \frac{(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})^*\Sigma_n^{-1}(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})\lambda^2\delta^2}{n(1-e^{-\lambda\delta})^2}. \tag{4.42}$$

Plugging (4.42) to $\frac{\partial l}{\partial \lambda}$ yields

$$\frac{\partial l}{\partial \lambda} = \frac{n(e^{-2\lambda\delta}S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}}) - e^{-\lambda\delta}S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(Y(\delta)^{\mathcal{D}}))}{(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})^*\Sigma_n^{-1}(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\lambda\delta})}.$$

Thus the solution of $\frac{\partial l}{\partial \lambda} = 0$ is that

$$\hat{\lambda}_T(\delta) = \frac{-1}{\delta}log\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(Y(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})}. \tag{4.43}$$

By replacing λ with $\hat{\lambda}_T(\delta)$ in (4.42), $\hat{\sigma}_T(\delta)$ becomes

$$\begin{aligned} &\hat{\sigma}_T^2(\delta) \\ &= \frac{\hat{\lambda}_T^2(\delta)\delta^2}{(1-e^{-\hat{\lambda}_T(\delta)\delta})^2}\frac{(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\hat{\lambda}_T(\delta)\delta})^*\Sigma_n^{-1}(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\hat{\lambda}_T(\delta)\delta})}{n} \\ &= \frac{1}{n(1-S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}\Delta S_{-1}^n(Y(\delta)^{\mathcal{D}}))^2}\left(log\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(Y(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})}\right)^2 \\ &\times S_0^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(Y(\delta)^{\mathcal{D}}) - 2(S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}\Delta S_{-1}^n(Y(\delta)^{\mathcal{D}}))S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}}) \\ &+ (S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}\Delta S_{-1}^n(Y(\delta)^{\mathcal{D}}))^2S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}}). \end{aligned} \tag{4.44}$$

As we can see that the expression for $\hat{\sigma}_T^2(\delta)$ is very messy, and throughout the thesis, we express $\hat{\sigma}_T^2(\delta)$ by the first line of the above equation.

4.5.2 Formulation for the consistency of the discrete model

Let λ_0 and σ_0 denote the parameters induced by the null hypothesis and P_0 the probability measure under λ_0 and σ_0 . We first prove the weak convergence for $\hat{\lambda}_T(\delta)$, that is to prove the error between the estimator $\hat{\lambda}_T(\delta)$ and λ_0 will vanish as $n \to \infty$ in probability. The first step is to find the error between $\hat{\lambda}_T(\delta)$ and λ_0, and hence, we need to reformulate the estimator. By applying Taylor's expansion theory (A.1.1) to (4.43) at $e^{-\lambda_0\delta}$, we have,

$$\begin{aligned}
\hat{\lambda}_T(\delta) &= \frac{-1}{\delta} log \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(Y(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \\
&= \frac{-1}{\delta} log \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} (\sigma_0 \frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta} S_0^n(\Delta X(\delta)^{\mathcal{D}}) + S_{-1}^n(Y(\delta)^{\mathcal{D}}) e^{-\lambda_0\delta})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \\
&= \frac{-1}{\delta} log (e^{-\lambda_0\delta} + \sigma_0 \frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta} \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}) \\
&= \lambda_0 - \frac{\sigma_0 \frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta} e^{\lambda_0\delta}}{\delta} \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \\
&+ \frac{1}{\delta} (\sigma_0 \frac{1-e^{-\lambda_0\delta}}{\xi\lambda_0\delta})^2 (\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})})^2,
\end{aligned} \tag{4.45}$$

where $\xi \in [e^{-\lambda_0\delta}, e^{-\lambda_0\delta} + \sigma_0 \frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta} \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}]$. (4.45) implies that, if we have

$$\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \xrightarrow{P_0} 0, \tag{4.46}$$

the weak convergence of $\hat{\lambda}_T(\delta)$ follows immediately.

Remark Remember that δ is fixed, and thus the coefficient $\frac{\sigma_0 \frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta} e^{\lambda_0\delta}}{\delta}$ is always finite. However, in the limiting case, we have to consider the coefficient, because

$$\frac{\sigma_0 \frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta} e^{\lambda_0\delta}}{\delta} \sim O(\frac{1}{\delta}) \xrightarrow{\delta \to 0} \infty. \tag{4.47}$$

We would like to apply the invariant distribution assumption for the proof, which assumes that the process follows the same distribution from $(-\infty, T]$, and thus we

find another expression for (4.46). Recall from (4.37), we have under P_0

$$Y(\delta)_{t_i} = \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\Delta X(\delta)_{t_i} + Y(\delta)_{t_{i-1}}e^{-\lambda_0\delta}.$$

By iteration, one can derive that

$$\begin{aligned} Y(\delta)_{t_i} &= e^{-i\lambda_0\delta}y(\delta)_0 + \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\sum_{k=0}^{i-1} e^{-k\lambda_0\delta}\Delta X(\delta)_{t_{i-k}} \\ &= \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\sum_{k=0}^{\infty} e^{-k\lambda_0\delta}\Delta X(\delta)_{t_{i-k}}. \end{aligned} \tag{4.48}$$

As we have the expression for $Y(\delta)_{t_i}$, we can derive the expression for $S^n_{-1}(Y(\delta)^{\mathcal{D}})$ as follows,

$$\begin{aligned} &S^n_{-1}(Y(\delta)^{\mathcal{D}}) = [Y(\delta)_{t_{n-1}}, Y(\delta)_{t_{n-2}}, ..., Y(\delta)_{t_0}]^* \\ &= \frac{\sigma(1-e^{-\lambda_0\delta})}{\lambda_0\delta}[\sum_{k=0}^{\infty} e^{-k\lambda_0\delta}\Delta X(\delta)_{t_{n-1-k}}, \sum_{k=0}^{\infty} e^{-k\lambda\delta}\Delta X(\delta)_{t_{n-2-k}}, ..., \sum_{k=0}^{\infty} e^{-k\lambda\delta}\Delta X(\delta)_{t_{0-k}}]^* \\ &= \frac{\sigma(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\sum_{k=0}^{\infty} e^{-k\lambda_0\delta}[\Delta X(\delta)_{t_{n-1-k}}, \Delta X(\delta)_{t_{n-2-k}}, ..., \Delta X(\delta)_{t_{0-k}}]^* \\ &= \frac{\sigma(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\sum_{k=0}^{\infty} e^{-k\lambda_0\delta}S^n_{-(k+1)}(\Delta X(\delta)^{\mathcal{D}}). \end{aligned} \tag{4.49}$$

We would like to make an observation on $S^n_{-k}(\Delta X(\delta)^{\mathcal{D}})$. We consider the eigenvalue decomposition $\Sigma_n = U\Lambda U^*$, where U is orthonormal, and thus $\Sigma^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^*$ is symmetric. Then we have $\Sigma_n^{\frac{1}{2}}\Sigma_n^{\frac{1}{2}} = \Sigma_n$. By the same virtue, we define $\Sigma^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}}U^*$, and thus $\Sigma_n^{-\frac{1}{2}}\Sigma_n^{-\frac{1}{2}} = \Sigma_n^{-1}$. Therefore, $I_n = \Sigma_n^{\frac{1}{2}}\Sigma_n^{-1}\Sigma_n^{\frac{1}{2}}$. First we make a notation

$$Z^n_{-k} := \Sigma_n^{-\frac{1}{2}}S^n_{-k}(\Delta X(\delta)^{\mathcal{D}}), \tag{4.50}$$

and hence Z^n_{-k} follows the standard multivariate normal distribution

$$\begin{aligned} f_{Z^n_{-k}}(z^n_{-k}) &= |\Sigma_n^{\frac{1}{2}}| f_{S^n_{-k}(\Delta X(\delta)^{\mathcal{D}})}(\Sigma_n^{\frac{1}{2}}z^n_{-k}) \\ &= \frac{|\Sigma_n^{\frac{1}{2}}|}{\sqrt{(2\pi)^n|\Sigma_n|}}e^{-\frac{1}{2}z^{n*}_{-k}\Sigma_n^{\frac{1}{2}}\Sigma_n^{-1}\Sigma_n^{\frac{1}{2}}z^n_{-k}} = \frac{1}{\sqrt{(2\pi)^n}}e^{-\frac{1}{2}z^{n*}_{-k}z^n_{-k}}. \end{aligned}$$

Let $< ., . >$ be the inner product of two vectors, and then the numerator of (4.46) is as follows,

$$
\begin{aligned}
&S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}}) = (S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-\frac{1}{2}})(\Sigma_n^{-\frac{1}{2}}S_0^n(\Delta X(\delta)^{\mathcal{D}})) \\
&= \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\left(\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}S_{-(k+1)}(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-\frac{1}{2}}\right)\left(\Sigma_n^{-\frac{1}{2}}S_0^n(\Delta X(\delta)^{\mathcal{D}})\right) \\
&= \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta}(\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}(\Sigma_n^{-\frac{1}{2}}S_{-(k+1)}(\Delta X(\delta)^{\mathcal{D}}))^*Z_0^n) \\
&= \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta} < \sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-(k+1)}^n, Z_0^n >
\end{aligned}
$$

and the denominator under P_0 is,

$$
\begin{aligned}
S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}}) &= (S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-\frac{1}{2}})(\Sigma_n^{-\frac{1}{2}}S_{-1}^n(Y(\delta)^{\mathcal{D}})) \\
&= (\frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta})^2(\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^{n*})(\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^n) \\
&= (\frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta})^2 < \sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^n > .
\end{aligned}
$$

Therefore (4.46) becomes as follows under P_0,

$$
\begin{aligned}
&\hat{\lambda}_T(\delta) - \lambda_0 \\
&= -\frac{e^{\lambda_0\delta}}{\delta}\sigma_0\frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta}\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})} \\
&+ \frac{e^{2\lambda_0\delta}}{\delta}(\sigma_0\frac{1-e^{-\lambda_0\delta}}{\xi\lambda_0\delta}\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})})^2 \\
&= -\frac{e^{\lambda_0\delta}}{\delta}\sigma_0\frac{1-e^{-\lambda_0\delta}}{\lambda_0\delta}\frac{< \sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^n, Z_0^n >}{< \sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^n >} \\
&+ \frac{e^{2\lambda_0\delta}}{\delta}(\sigma_0\frac{1-e^{-\lambda_0\delta}}{\xi\lambda_0\delta}\frac{< \sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^n, Z_0^n >}{< \sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z_{-k-1}^n >})^2 \\
&= -\frac{e^{\lambda_0\delta}}{\delta}\frac{N_n}{D_n} + \frac{e^{2\lambda_0\delta}}{\delta\xi^2}(\frac{N_n}{D_n})^2,
\end{aligned}
$$

where $< \sum_{k=0}^{\infty} e^{-k\lambda_0\delta} Z_{-k-1}^n, Z_0^n >$ is denoted by N_n, and $< \sum_{k=0}^{\infty} e^{-k\lambda_0\delta} Z_{-k-1}^n >$ is denoted by D_n.

Thus, the consistence of the parameter is formulated as follows,

$$|\hat{\lambda}_T(\delta) - \lambda_0| \xrightarrow{P_0} 0 \iff \frac{N_n}{D_n} \xrightarrow{P_0} 0 \tag{4.51}$$

and we proceed to prove that $\frac{N_n}{D_n} \xrightarrow{P_0} 0$ in the next section.

4.5.3 Consistency for the drift coefficient of the discrete model

In this section, we focus on the proof for the consistency of the drift coefficient, and by Slutsky's theorem (for converge in probability)(A.1.2), we separate the proof of (4.51) into two steps.

1. We prove that $\frac{N_n}{n} \to 0$, as $n \to \infty$ in P_0.
2. We prove that $\frac{D_n}{n} \to \frac{1}{1-e^{-2\lambda_0\delta}}$ as $n \to \infty$ in P_0.

We consider the numerator first and we prove the following lemma,

Lemma 4.5.1. $\frac{N_n}{n} \xrightarrow{P_0} 0$ *as* $n \to \infty$ *with* $\delta > 0$ *fixed.*

Proof. Since Σ_n^{-1} is symmetric, we define

$$G_k(n) := S_{-d}^n(\Delta X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-l}^n(\Delta X(\delta)^{\mathcal{D}}) =< Z_{-d}^n, Z_{-l}^n >,$$

where $|d-l|=k$. By Markov's inequality, we have

$$P_0(|\frac{N_n}{n}|>\epsilon)=P_0(|\frac{<\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z^n_{-k-1},Z^n_0>}{n}|>\epsilon)\leq\frac{1}{n^2\epsilon^2}\mathbb{E}\left(<\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}Z^n_{-k-1},Z^n_0>\right)^2$$
$$=\frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty}e^{-k\lambda_0\delta}<Z^n_{-k-1},Z^n_0>\right)^2=\frac{e^{\lambda_0\delta}}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=1}^{\infty}e^{-k\lambda_0\delta}G_k(n)\right)^2$$
$$\sim O(\frac{1}{n^2}\frac{e^{-\lambda_0\delta}}{\epsilon^2(1-e^{-\lambda_0\delta})}\sum_{k=1}^{\infty}e^{-k\lambda_0\delta}\mathbb{E}G_k(n)^2)$$
$$\sim O(\frac{1}{n^2}\frac{e^{-\lambda_0\delta}}{\epsilon^2(1-e^{-\lambda_0\delta})}\sum_{k=1}^{\infty}e^{-k\lambda_0\delta}\{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d}(j))\mathbb{E}(Z^n_{-d-k}(i)Z^n_{-d-k}(j))$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(i))\mathbb{E}(Z^n_{-d}(j)Z^n_{-d-k}(j))$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(j))\mathbb{E}(Z^n_{-d}(j)Z^n_{-d-k}(i))\}), \tag{4.52}$$

and for the details of the above calculation, please refer to Appendix A.2.1.
For the first summand, we have almost surely,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d}(j))\mathbb{E}(Z^n_{-d-k}(i)Z^n_{-d-k}(j))$$
$$=\sum_{i=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d}(i))\mathbb{E}(Z^n_{-d-k}(i)Z^n_{-d-k}(i))\sim O(n),$$

and

$$\frac{\sum_{k=1}^{\infty}e^{-k\lambda_0\delta}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d}(j))\mathbb{E}(Z^n_{-d-k}(i)Z^n_{-d-k}(j))}{n^2}\sim O(n^{-1}\delta^{-1}). \tag{4.53}$$

We define a matrix V_k as follows,

$$V_k:=\mathbb{E}(S^n_{-d}(\Delta X^{\mathcal{D}})S^n_{-d-k}(\Delta X^{\mathcal{D}})^*)=\mathbb{E}(S^n_0(\Delta X^{\mathcal{D}})S^n_{-k}(\Delta X^{\mathcal{D}})^*),$$

$trace(A)$ is the trace of a matrix A, and $cov(A,B):=\mathbb{E}(AB^*)$ is the covariance matrix between random vector A and B.

For the second summand, we denote the following

$$\begin{aligned} A_1^k(n) &:= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(i))\mathbb{E}(Z^n_{-d}(j)Z^n_{-d-k}(j)) \\ &= \left(\sum_{i=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(i))\right)^2 = \left(\sum_{i=1}^{n} cov(Z^n_{-d}, Z^n_{-d-k})(i,i)\right)^2 \\ &= \left(\sum_{i=1}^{n} cov(\Sigma^{-\frac{1}{2}}S_d(S_0^n(\Delta X(\delta)^{\mathcal{D}})), \Sigma^{-\frac{1}{2}}S_{d+k}(S_0^n(\Delta X(\delta)^{\mathcal{D}})))(i,i)\right)^2 \\ &= trace(\Sigma_n^{-\frac{1}{2}}V_k\Sigma_n^{-\frac{1}{2}*})^2 = trace(V_k\Sigma_n^{-1})^2, \end{aligned} \tag{4.54}$$

and for the third summand, we denote the following

$$\begin{aligned} A_2^k(n) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(j))\mathbb{E}(Z^n_{-d}(j)Z^n_{-d-k}(i)) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} cov(Z^n_{-d}, Z^n_{-d-k})(i,j)cov(Z^n_{-d}, Z^n_{-d-k})(j,i) \\ &= trace(\Sigma_n^{-\frac{1}{2}}V_k\Sigma_n^{-\frac{1}{2}*}\Sigma_n^{-\frac{1}{2}}V_k\Sigma_n^{-\frac{1}{2}*}) = trace(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}). \end{aligned} \tag{4.55}$$

The expression requires us to study the trace of the covariance matrices $V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}$ and $V_k\Sigma_n^{-1}V_k$, which can be found in Appendix A.2.2, and from which two lemmas are derived.

Lemma 4.5.2. $\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A_k^1(n)}{n^2} \sim O(n^{-2}\delta^{-3})$.

Lemma 4.5.3. $\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A_k^2(n)}{n^2} \sim O(n^{-2}\delta^{-3})$.

Th proof of the two lemmas can be found in Appendix A.2.3.
From equation (4.53), Lemma 4.5.2 and Lemma 4.5.3, the result follows immediately. □

Next, we prove the second step for the Theorem 4.5.1.

Lemma 4.5.4. $\frac{D_n}{n} \xrightarrow{P_0} \frac{1}{1-e^{-2\lambda_0\delta}}$ *as* $n \to \infty$ *with* δ *fixed.*

Proof. First, we expand D_n as follows,

$$D_n = < \sum_{k=0}^{\infty} e^{-k\lambda_0\delta}Z^n_{-k-1} > = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} e^{-(k+l)\lambda_0\delta} < Z^n_{-l-1}, Z^n_{-k-1} >$$

$$= \sum_{k=0}^{\infty} e^{-2k\lambda_0\delta} < Z_{-k-1}^n > +2\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} < Z_{-l-1}^n, Z_{-k-1}^n >$$
$$= \sum_{k=0}^{\infty} e^{-2k\lambda_0\delta} < Z_{-k-1}^n > +2\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n).$$

By Markov's inequality, we have

$$\begin{aligned}
&P_0(|\frac{D_n}{n} - \frac{1}{1-e^{-2\lambda_0\delta}}| > \epsilon) \leq \frac{1}{n^2\epsilon^2}\mathbb{E}(D_n - \frac{n}{1-e^{-2\lambda_0\delta}})^2 \\
&= \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n) + 2\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n)\right)^2 \\
&= \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n)\right)^2 + \frac{4}{n^2}\mathbb{E}\left(\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n)\right)^2 \\
&+ \frac{4}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n)\right)\left(\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n)\right) \\
&= S_1(n) + S_2(n) + A(n),
\end{aligned}$$

where

$$S_1(n) := \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n) + 2\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n)\right)^2,$$

$$S_2(n) := \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n)\right)^2 + \frac{4}{n^2}\mathbb{E}\left(\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n)\right)^2,$$

and

$$A(n) := \frac{4}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n)\right)\left(\sum_{l=0}^{\infty}\sum_{k>l}^{\infty} e^{-(k+l)\lambda_0\delta} G_{k-l}(n)\right).$$

As $A(n)$ is controlled by $S_1(n)$ and $S_2(n)$, we only consider $S_1(n)$ and $S_2(n)$. For $S_1(n)$, we have

$$S_1(n) = \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}(< Z_{-k-1}^n > -n)\right)^2 = \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-2k\lambda_0\delta}\sum_{i=1}^{n}((Z_{-k-1}^n(i))^2 - 1)\right)^2$$

$$
\begin{aligned}
&= \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-4k\lambda_0\delta}\left(\sum_{i=1}^{n}(Z^n_{-k-1}(i)^2-1)\right)^2\right)\\
&+ \frac{2}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty}\sum_{l>k}^{\infty} e^{-2(k+l)\lambda_0\delta}\sum_{i=1}^{n}\sum_{j=0}^{n}(Z^n_{-k-1}(i)^2-1)(Z^n_{-l-1}(j)^2-1)\right)\\
&= S_{1,1}(n) + S_{1,2}(n),
\end{aligned}
$$

where

$$
S_{1,1}(n) := \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-4k\lambda_0\delta}\left(\sum_{i=1}^{n}(Z^n_{-k-1}(i)^2-1)\right)^2\right),
$$

and

$$
S_{1,2}(n) := \frac{2}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty}\sum_{l>k}^{\infty} e^{-2(k+l)\lambda_0\delta}\sum_{i=1}^{n}\sum_{j=0}^{n}(Z^n_{-k-1}(i)^2-1)(Z^n_{-l-1}(j)^2-1)\right).
$$

First, for any $k, l \in \mathbb{N}$, and $i, j \in N$, we have, from Isserli's Theorem (A.1.6),

$$
\begin{aligned}
&\mathbb{E}(Z^n_{-k-1}(i)^2-1)(Z^n_{-l-1}(j)^2-1) = \mathbb{E}(Z^n_{-k-1}(i)^2 Z^n_{-l-1}(j)^2 - Z^n_{-k-1}(i)^2 - Z^n_{-k-1}(i)^2 + 1)\\
&= \mathbb{E}(Z^n_{-k-1}(i)^2 Z^n_{-l-1}(j)^2) - 1\\
&= \mathbb{E}(Z^n_{-k-1}(i)^2)\mathbb{E}(Z^n_{-l-1}(j)^2) + 2(\mathbb{E}(Z^n_{-k-1}(i) Z^n_{-l-1}(j)))^2 - 1\\
&= 2(\mathbb{E}(Z^n_{-k-1}(i) Z^n_{-l-1}(j)))^2.
\end{aligned}
\tag{4.56}
$$

For $S_{1,1}(n)$, from (4.56), we have

$$
\begin{aligned}
&\frac{1}{n^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-4k\lambda_0\delta}\left(\sum_{i=1}^{n}(Z^n_{-k-1}(i)^2-1)\right)^2\right) = \frac{1}{n^2}\sum_{k=0}^{\infty} e^{-4k\lambda_0\delta}\mathbb{E}\left(\sum_{i=1}^{n}(Z^n_{-k-1}(i)^2-1)\right)^2\\
&= \frac{1}{n^2}\sum_{k=0}^{\infty} e^{-4k\lambda_0\delta}\left(\sum_{i=1}^{n}\mathbb{E}(Z^n_{-k-1}(i)^2-1)^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n}\mathbb{E}(Z^n_{-k-1}(i)^2-1)(Z^n_{-k-1}(j)^2-1)\right)\\
&= \frac{2}{n(1-e^{-4\lambda_0\delta})},
\end{aligned}
\tag{4.57}
$$

and for $S_{1,2}(n)$, by (4.56)

$$\begin{aligned}
S_{1,2}(n) &= \frac{2}{n^2}\mathbb{E}\left(\sum_{k=0}^{\infty}\sum_{l>k}^{\infty} e^{-2(k+l)\lambda_0\delta}\sum_{i=1}^{n}\sum_{j=0}^{n}(Z^n_{-k-1}(i)^2-1)(Z^n_{-l-1}(j)^2-1)\right)\\
&= \frac{2}{n^2}\left(\sum_{k=0}^{\infty}\sum_{l>k}^{\infty} e^{-2(k+l)\lambda_0\delta}2\sum_{i=1}^{n}\sum_{j=0}^{n}\mathbb{E}(Z^n_{-k-1}(i)Z^n_{-l-1}(j))^2\right)\\
&= \frac{4}{n^2}\left(\sum_{k=0}^{\infty}\sum_{d=1}^{\infty} e^{-(2k+d)\lambda_0\delta}\sum_{i=1}^{n}\sum_{j=0}^{n}C_d(i,j)^2\right)\\
&= \frac{4}{n^2}\left(\sum_{k=0}^{\infty}\sum_{d=1}^{\infty} e^{-(2k+d)\lambda_0\delta}\|C_d\|_F^2\right)\\
&\sim O(\frac{1}{n}\frac{1}{(1-e^{-\lambda_0\delta})^3}),
\end{aligned} \tag{4.58}$$

where

$$\begin{aligned}
C_d &:= \mathbb{E}(Z^n_{-k}Z^{n*}_{-k-d}) = \mathbb{E}(\Sigma_n^{-\frac{1}{2}}S^n_{-k}(\Delta X(\delta)^{\mathcal{D}})(\Sigma_n^{-\frac{1}{2}}S^n_{-k-d}(\Delta X(\delta)^{\mathcal{D}}))^*)\\
&= \Sigma_n^{-\frac{1}{2}}\mathbb{E}(S^n_{-k}(\Delta X(\delta)^{\mathcal{D}})S^n_{-k-d}(\Delta X(\delta)^{\mathcal{D}})^*)\Sigma_n^{-\frac{1}{2}*}\\
&= \Sigma_n^{-\frac{1}{2}}V_d\Sigma_n^{-\frac{1}{2}*} = \Sigma_n^{-\frac{1}{2}}V_d\Sigma_n^{-\frac{1}{2}},
\end{aligned}$$

and $\|C_d\|_F$ is the Frobenius norm [44] which is defined as

$$\|C_d\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=0}^{n}C_d(i,j)^2}.$$

The calculation for (4.58) is in Appendix A.2.4.
For $S_2(n)$, from Theorem 4.5.1, we have

$$\frac{1}{n^2}\mathbb{E}\left(\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}G_d(n)\right)^2 \sim O(n^{-2}\delta^{-3}),$$

and thus,

$$\begin{aligned} S_2(n) &= \frac{4}{n^2}\mathbb{E}\left(\sum_{l=1}^{\infty} e^{-2l\lambda_0\delta}\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}G_d(n)\right)^2 \\ &= \frac{1}{(1-e^{-2\lambda_0\delta})^2}\frac{4}{n^2}\mathbb{E}\left(\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}G_d(n)\right)^2 \\ &\sim O(n^{-2}\delta^{-5}). \end{aligned} \tag{4.59}$$

Therefore, by (4.57), (4.58), and (4.59), the result follows immediately. □

Therefore, by Lemma 4.5.1 and Lemma 4.5.4, the following theorem follows immediately by Slutsky's theorem (A.1.2),

Theorem 4.5.1.

$$\lim_{T\to\infty} \hat{\lambda}_T(\delta) \xrightarrow{P_0} \lambda_0,$$

as $T = n\delta \to \infty$ with δ fixed.

Remark We would like to make a remark on the conditions for the weak consistency with respect to δ and n. At this stage, δ is fixed, but the limiting case requires $\delta \to 0$. The weak convergence of the limiting case depends on Lemma 4.5.1 and Lemma 4.5.4, some conditions have to be placed with respect to δ and n, for the limiting case. For Lemma 4.5.1, we need

$$n^{-1}\delta^{-1} \to 0, \quad \text{and} \quad n^{-2}\delta^{-3} \to 0,$$

and for Lemma 4.5.4, from (4.57), (4.58), and (4.59), we need

$$n^{-1}\delta^{-3} \to 0.$$

For both of the above conditions to satisfy, we require that

$$\delta n^{\frac{1}{3}} \to \infty. \tag{4.60}$$

The result is useful for the consistency of the limiting model.

4.5.4 Consistency for the diffusion coefficient of the discrete model

We now proceed to prove the weak consistency of $\hat{\sigma}_T^2(\delta)$.

Theorem 4.5.2.

$$\lim_{T\to\infty} \hat{\sigma}_T^2(\delta) \xrightarrow{P_0} \sigma_0^2,$$

as $T = n\delta \to \infty$ with δ fixed.

Proof. Recall from (4.42) that

$$\hat{\sigma}_T^2(\delta) = \frac{\hat{\lambda}_T^2(\delta)\delta^2}{(1-e^{-\hat{\lambda}_T(\delta)\delta})^2} \frac{(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\hat{\lambda}_T(\delta)\delta})^* \Sigma_n^{-1} (S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\hat{\lambda}_T(\delta)\delta})}{n}.$$

The Continuous Mapping Theorem (A.1.9) implies that

$$\frac{\hat{\lambda}_T^2(\delta)\delta^2}{(1-e^{-\hat{\lambda}_T(\delta)\delta})^2} \xrightarrow{P_0} \frac{\lambda_0^2\delta^2}{(1-e^{-\lambda_0\delta})^2}. \tag{4.61}$$

For the other part, we have

$$\begin{aligned}
&\frac{\left(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\hat{\lambda}_T(\delta)\delta}\right)^* \Sigma_n^{-1} \left(S_0^n(Y(\delta)^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})e^{-\hat{\lambda}_T(\delta)\delta}\right)}{n} \\
&= \frac{\left(S_{-1}^n(Y(\delta)^{\mathcal{D}})(e^{-\lambda_0\delta} - e^{-\hat{\lambda}_T(\delta)\delta}) + \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta} S_0^n(\Delta X(\delta)^{\mathcal{D}})\right)^* \Sigma_n^{-1}}{n} \quad \dots \\
&\times \left(S_{-1}^n(Y(\delta)^{\mathcal{D}})(e^{-\lambda_0\delta} - e^{-\hat{\lambda}_T(\delta)\delta}) + \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta} S_0^n(\Delta X(\delta)^{\mathcal{D}})\right) \\
&= \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n} (e^{-\hat{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})^2 \\
&+ \frac{\sigma_0^2(1-e^{-\lambda_0\delta})^2}{\lambda_0^2\delta^2} \frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{n} \\
&+ 2(e^{-\hat{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta}) \frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta} \frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n} \\
&= A_1 + A_2 + A_3,
\end{aligned} \tag{4.62}$$

where

$$A_1 := \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n} (e^{-\hat{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})^2,$$

$$A_2 := \frac{\sigma_0^2(1-e^{-\lambda_0\delta})^2}{\lambda_0^2\delta^2} \frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})}{n},$$

and

$$A_3 := 2(e^{-\hat{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})\frac{\sigma_0(1-e^{-\lambda_0\delta})}{\lambda_0\delta}\frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n}.$$

For A_1, we have $\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n} \xrightarrow{P_0} \frac{1}{1-e^{-2\lambda_0\delta}}$, and $e^{-\hat{\lambda}_T(\delta)\delta} \xrightarrow{P_0} e^{-\lambda_0\delta}$, which yields that $(e^{-\hat{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})^2 \xrightarrow{P_0} 0$ by the Continuous Mapping Theorem. Therefore $A_1 \xrightarrow{P_0} 0$.

For A_2, by the law of large number,

$$\frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})}{n} = \frac{< Z_0^n, Z_0^n >}{n} \xrightarrow{P_0} 1,$$

where recall from (4.50), Z_0^n is a series of independent standard normal random variables of length n. Therefore, we have $A_2 \xrightarrow{P_0} \frac{\sigma_0^2(1-e^{-\lambda_0\delta})^2}{\lambda_0^2\delta^2}$.

Since $|\frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n}| \xrightarrow{P_0} 0$, we have $A_3 \xrightarrow{P_0} 0$. Therefore, we have

$$(4.62) \xrightarrow{P_0} \frac{\sigma_0^2(1-e^{-\lambda_0\delta})^2}{\lambda_0^2\delta^2}.$$

Therefore by (4.61) and (4.62), we have $\hat{\sigma}_T^2(\delta) \xrightarrow{P_0} \sigma_0^2$, as $n \to \infty$ with fixed δ. □

4.5.5 Formulation for the consistency of the limiting model

In this section, we prove the weak consistency of the estimators for the limiting case. Recall from (4.36), $\mathcal{D} = \{t_n, t_{n-1}, ..., t_0, t_{-1}, ...\}$ where the time between observations is $t_i - t_{i-1} = \delta$, and n is the number of observations. Let $Y^{\mathcal{D}} \in \mathbb{R}^\infty$ be an infinite vector as follows,

$$Y^{\mathcal{D}} = [Y_{t_n}, Y_{t_{n-1}}, ..., Y_{t_{-1}}, ...]^*,$$

and

$$\Delta Y^{\mathcal{D}} = [Y_{t_n} - Y_{t_{n-1}}, Y_{t_{n-1}} - Y_{t_{n-2}}, ..., Y_{t_{-1}} - Y_{t_{-2}}, ...]^*.$$

The MLEs for λ and σ are constructed by applying the observations from the limiting model to the MLEs (4.43) and (4.44) derived from the discrete model. Let $\tilde{\lambda}_T(\delta)$ and $\tilde{\sigma}_T(\delta)$ denote the estimators for the limiting case, and we have

$$\tilde{\lambda}_T(\delta) = \frac{-1}{\delta} log \frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(Y^{\mathcal{D}})}{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})}, \tag{4.63}$$

and

$$\begin{aligned}
&\tilde{\sigma}_T^2(\delta) \\
&= \frac{1}{n(1 - S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta Y^{\mathcal{D}}))^2} \left(log \frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(Y^{\mathcal{D}})}{S_{-1}^n(Y^{\mathcal{D}}) \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})} \right)^2 \\
&\times S_0^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(Y^{\mathcal{D}}) - 2(S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(\Delta Y^{\mathcal{D}})) S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}}) \\
&+ (S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(\Delta Y^{\mathcal{D}}))^2 S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}}).
\end{aligned} \tag{4.64}$$

The resulting MLEs have two sources of error, and one is the data-model mismatching error, since the log-likelihood function is derived from the discrete model instead of the limiting model. The other error is due to the missing data. Since the observation is discrete data points, the information between each observation point is missing. The two sources of error are expected to vanish as $\delta \to 0$ and $n \to \infty$. Thus we would like to prove the consistency when $\delta \to 0$, which is different from the discrete model where $\delta > 0$ is fixed. We first consider the drift coefficient λ. First, we would like to find the relationship between $Y^{\mathcal{D}}$ and $Y(\delta)^{\mathcal{D}}$. By [12], Y_{t_i} admits the following expression,

$$\begin{aligned}
Y_{t_{i+1}} &= \sigma_0 \int_{-\infty}^{t_{i+1}} e^{-(t_{i+1}-s)\lambda_0} dX_s = \sigma_0 e^{-\delta\lambda_0} \left(\int_{-\infty}^{t_i} e^{-(t_i-s)\lambda_0} dX_s + \int_{t_i}^{t_{i+1}} e^{-(t_i-s)\lambda_0} dX_s \right) \\
&= e^{-\delta\lambda_0} Y_{t_i} + \sigma_0 \int_0^{\delta} e^{-u\lambda_0} dX_{t_{i+1}-u},
\end{aligned} \tag{4.65}$$

where $(t_{i+1} - s) = u$. We denote $\int_0^{\delta} e^{-u\lambda_0} dX_{t_{i+1}-u}$ by $\Delta^{\lambda} X_{t_{i+1}}$, and the infinite random vector $\Delta^{\lambda} X^{\mathcal{D}}$ by

$$\Delta^{\lambda} X^{\mathcal{D}} = [\Delta^{\lambda} X_{t_n}, \Delta^{\lambda} X_{t_{n-1}}, ..., \Delta^{\lambda} X_{t_{-1}}, ...]^*.$$

The distance between $\tilde{\lambda}_T(\delta)$ and the true parameter λ_0 is

$$\begin{aligned}
|\tilde{\lambda}_T(\delta) - \lambda_0| &= |-\frac{1}{\delta} log \frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_0(Y^{\mathcal{D}})}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})} - \lambda_0| \\
&= |-\frac{1}{\delta} log \frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} (e^{-\lambda_0 \delta} S^n_{-1}(Y^{\mathcal{D}}) + \sigma_0 S^n_0(\Delta^\lambda X^{\mathcal{D}}))}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})} - \lambda_0| \\
&= |-\frac{1}{\delta} log \left(e^{-\lambda_0 \delta} + \sigma_0 \frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_0(\Delta^\lambda X^{\mathcal{D}})}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})} \right) - \lambda_0| \\
&= |\sigma_0 \frac{-1}{\delta} \frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_0(\Delta^\lambda X^{\mathcal{D}})}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})} + \frac{\xi^{-2}}{\delta} (\sigma_0 \frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_0(\Delta^\lambda X^{\mathcal{D}})}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})})^2|
\end{aligned},$$

where $\xi \in [e^{-\lambda_0 \delta}, e^{-\lambda_0 \delta} + \sigma_0 e^{-\lambda_0 \delta} \frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_0(\Delta^\lambda X^{\mathcal{D}})}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})}]$. The final step is due to the Taylor expansion (A.1.1) at $e^{-\lambda_0 \delta}$. Therefore,

$$\begin{aligned}
&|\tilde{\lambda}_T(\delta) - \lambda_0| \xrightarrow{P_0} 0 \iff \\
&\frac{1}{\delta} |\frac{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_0(\Delta^\lambda X^{\mathcal{D}})}{S^n_{-1}(Y^{\mathcal{D}})^* \Sigma_n^{-1} S^n_{-1}(Y^{\mathcal{D}})}| \xrightarrow{P_0} 0.
\end{aligned} \tag{4.66}$$

We first make an observation which gives another form of expression of $Y_{t_{i+1}}$.

Observation 4.5.1.

$$Y_{t_{i+1}} = \sigma_0 \sum_{j=0}^{\infty} e^{-j\delta\lambda_0} \left(\Delta X(\delta)_{t_{i+1-j}} - \lambda_0 I_{t_{i-j+1}}(\delta) + \frac{\lambda_0^2}{2} R_{t_{i-j+1}}(\delta) \right),$$

where $I(\delta)_{t_{i-j+1}} = \int_0^\delta u dX_{t_{i-j+1}-u}$, *and* $R(\delta)_{t_{i-j+1}} = \int_0^\delta f(u) dX_{t_{i-j+1}-u}$ *with* $f(u) = e^{-\lambda_0 u} - 1 + \lambda_0 u \sim O(u^2)$, *the residual function of the Taylor expansion of* $e^{-\lambda_0 u}$ *at* $u = 0$.

Proof. From (4.65), we have

$$\begin{aligned}
Y_{t_{i+1}} &= \sigma_0 \sum_{j=0}^{\infty} e^{-j\delta\lambda_0} \Delta^\lambda X_{t_{i+1-j}} \\
&= \sigma_0 \sum_{j=0}^{\infty} e^{-j\delta\lambda_0} \int_0^\delta e^{-u\lambda_0} dX_{t_{i+1-j}-u}.
\end{aligned}$$

By applying the Taylor expansion to $e^{u\lambda_0}$ around 0, we have

$$\begin{aligned}\int_0^\delta e^{-u\lambda_0} dX_{t_{i+1-j}-u} &= \int_0^\delta (1-\lambda_0 u + \frac{\lambda_0^2}{2} f(u)) dX_{t_{i+1-j}-u} \\ &= \Delta X(\delta)_{t_{i+1-j}} - \lambda_0 I(\delta)_{t_{i-j+1}} + \frac{\lambda_0^2}{2} R(\delta)_{t_{i-j+1}},\end{aligned}$$

where $I(\delta)_{t_{i-j+1}} := \int_0^\delta u dX_{t_{i-j+1}-u}$ and $R(\delta)_{t_{i-j+1}} := \int_0^\delta f(u) dX_{t_{i+1-j}-u}$ with $f(u) \sim O(u^2)$ and $f(0) = 0$. Therefore, the result is proved. □

We define

$$I(\delta)^{\mathcal{D}} := [I(\delta)_{t_n}, I(\delta)_{t_{n-1}}, ..., I(\delta)_{t_{-1}}, ...]^* \in \mathbb{R}^\infty,$$

and

$$R(\delta)^{\mathcal{D}} := [R(\delta)_{t_n}, R(\delta)_{t_{n-1}}, ..., R(\delta)_{t_{-1}}, ...] \in \mathbb{R}^\infty.$$

Hence, we have

$$\Delta^\lambda X(\delta)^{\mathcal{D}} = \Delta X(\delta)^{\mathcal{D}} - \lambda_0 I(\delta)^{\mathcal{D}} + \frac{\lambda_0^2}{2} R(\delta)^{\mathcal{D}}, \tag{4.67}$$

and we can express $S_{-1}^n(Y^{\mathcal{D}})$ by the solution of the discrete model $S_{-1}^n(Y(\delta)^{\mathcal{D}})$ as follows,

$$\begin{aligned}S_{-1}^n(Y^{\mathcal{D}}) &= \sigma_0 \sum_{j=1}^\infty e^{(-j+1)\delta\lambda_0} S_{-j}^n(\Delta^\lambda X(\delta)^{\mathcal{D}}) \\ &= \sigma_0 \sum_{j=1}^\infty e^{(-j+1)\delta\lambda_0} S_{-j}^n \left(\Delta X(\delta)^{\mathcal{D}} - \lambda_0 I(\delta)^{\mathcal{D}} + \frac{\lambda_0^2}{2} R(\delta)^{\mathcal{D}} \right) \\ &= \frac{\lambda_0 \delta}{1-e^{-\delta\lambda_0}} S_{-1}^n(Y(\delta)^{\mathcal{D}}) - \sigma_0 \lambda_0 \sum_{j=1}^\infty e^{(-j+1)\delta\lambda_0} S_{-j}^n(I(\delta)^{\mathcal{D}}) \\ &+ \sigma_0 \frac{\lambda_0^2}{2} \sum_{j=1}^\infty e^{(-j+1)\delta\lambda_0} S_{-j}^n(R(\delta)^{\mathcal{D}}).\end{aligned} \tag{4.68}$$

Since $R(\delta)^{\mathcal{D}}$ is the residual, it is dominated by $I(\delta)^{\mathcal{D}}$, and we make an observation on the relationship between them.

Observation 4.5.2. *Consider $R(\delta)^{\mathcal{D}}$ and $I(\delta)^{\mathcal{D}}$. For any $t_i \in \mathcal{D}$, $\frac{|R(\delta)_{t_i}|}{|I(\delta)_{t_i}|} \sim O(\delta)$*

almost surely.

Proof. Recall that $I(\delta)_{t_i} = \int_0^\delta u dX_{t_i-u}$, and $R(\delta)_{t_i} = \int_0^\delta f(u) dX_{t_i-u}$ where $f(u) \sim O(u^2)$, and $f(0) = 0$. The integral admits Riemann integral limit by Young's theorem, and thus admits integral by parts. Therefore, we have

$$\begin{aligned}\frac{R(\delta)_{t_i}}{I(\delta)_{t_i}} &= \frac{\int_0^\delta f(u) dX_{t_i-u}}{\int_0^\delta u dX_{t_i-u}} \\ &= \frac{\int_0^\delta (X_{t_{i-1}} - X_{t_i-u}) df(u)}{\int_0^\delta (X_{t_{i-1}} - X_{t_i-u}) du} \sim O(\delta).\end{aligned}$$

□

4.5.6 Consistency for the drift coefficient of the limiting model

For (4.66), by Slustky's theorem (for convergence in probability)(A.1.2), we can separate the proof with respect to the nominator and the denominator. First, we consider the denominator, and we prove the following lemma

Lemma 4.5.5. $\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \xrightarrow{P_0} 1$, *for* $h > 0.5$, *with* $\delta \to 0$, *and* $\delta n^{\frac{1}{3}} \to \infty$; *for* $\frac{1}{6} < h < 0.5$, *with* $\delta \to 0$, $\delta n^{\frac{1}{3}} \to \infty$, *and* $n^{\frac{1}{2}-h}\delta \to 0$.

Proof. We define $Q(\delta) = S_{-1}^n(Y^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})$, and from (4.68), we have

$$\begin{aligned}Q(\delta) &= (\frac{\lambda_0 \delta}{1 - e^{-\delta\lambda_0}} - 1) S_{-1}^n(Y(\delta)^{\mathcal{D}}) - \sigma_0 \lambda_0 \sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} S_{-j}^n(I(\delta)^{\mathcal{D}}) \\ &\quad + \sigma_0 \frac{\lambda_0^2}{2} \sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} S_{-j}^n(R(\delta)^{\mathcal{D}}) \\ &= Q_1(\delta) + Q_2(\delta) + Q_3(\delta),\end{aligned} \tag{4.69}$$

where

$$Q_1(\delta) = (\frac{\lambda_0 \delta}{1 - e^{-\delta\lambda_0}} - 1) S_{-1}^n(Y(\delta)^{\mathcal{D}}),$$

$$Q_2(\delta) = \sigma_0 \lambda_0 \sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} S_{-j}^n(I(\delta)^{\mathcal{D}}),$$

and

$$Q_3(\delta) = \sigma_0 \frac{\lambda_0^2}{2} \sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} S_{-j}^n(R(\delta)^{\mathcal{D}}).$$

We have

$$\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \xrightarrow{P_0} 1 \iff$$
$$|\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}}) - S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}| \xrightarrow{P_0} 0,$$

that is

$$|\frac{2Q(\delta)^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} + \frac{Q(\delta)^* \Sigma_n^{-1} Q(\delta)}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}| \xrightarrow{P_0} 0.$$

Since $|\frac{Q(\delta)^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}| \leq \sqrt{\frac{Q(\delta)^* \Sigma_n^{-1} Q(\delta)}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}}$, by Chauchy's inequality (A.1.7), from Continuous Mapping Theorem (A.1.9), we only need to prove

$$\frac{Q(\delta)^* \Sigma_n^{-1} Q(\delta)}{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \xrightarrow{P_0} 0.$$

We can see that $Q_2(\delta)^* \Sigma_n^{-1} Q_2(\delta)$ dominates $Q_3(\delta)^* \Sigma_n^{-1} Q_3(\delta)$ by Observation 4.5.2. In addition, $Q_1(\delta)^* \Sigma_n^{-1} Q_1(\delta) \sim O(\delta^2) S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(Y(\delta)^{\mathcal{D}})$. Therefore, we only need to discuss $Q_2(\delta)^* \Sigma_n^{-1} Q_2(\delta)$.
From Lemma 4.5.4, we know that

$$\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})}{n} \to \frac{1}{1 - e^{-2\lambda_0\delta}},$$

under the condition that

$$\delta n^{\frac{1}{3}} \to \infty, \tag{4.70}$$

and by Slutsky's theorem (A.1.2), it is equivalent to prove

$$\frac{Q_2(\delta)^* \Sigma_n^{-1} Q_2(\delta)}{n(1 - e^{-2\lambda_0\delta})^{-1}} \xrightarrow{P_0} 0. \tag{4.71}$$

By Markov's inequality, we have

$$\mathbb{E}\left(\frac{Q_2(\delta)^* \Sigma_n^{-1} Q_2(\delta)}{n(1 - e^{-2\lambda_0\delta})^{-1}}\right) \to 0 \implies \frac{Q_2(\delta)^* \Sigma_n^{-1} Q_2(\delta)}{n(1 - e^{-2\lambda_0\delta})^{-1}} \xrightarrow{P_0} 0.$$

Thus we consider $\mathbb{E}Q_2(\delta)^*\Sigma_n^{-1}Q_2(\delta)$, and we have

$$\begin{aligned}\mathbb{E}Q_2(\delta)^*\Sigma_n^{-1}Q_2(\delta) &= \sigma_0^2\lambda_0^2\mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}S_{-j}^n(I(\delta)^{\mathcal{D}})\right)^*\Sigma_n^{-1}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}S_{-j}^n(I(\delta)^{\mathcal{D}})\right)\\ &\leq \sigma_0^2\lambda_0^2\|\Sigma_n^{-1}\|_2\mathbb{E}\|\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}S_{-j}^n(I(\delta)^{\mathcal{D}})\|_2^2\\ &= \sigma_0^2\lambda_0^2\|\Sigma_n^{-1}\|_2\sum_{i=1}^{n}\mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}S_{-j}^n(I(\delta)^{\mathcal{D}})(i)\right)^2,\end{aligned}$$

where recall that $\|.\|_2$ is the 2-norm of a matrix. Since increments of fractional Brownian motion are stationary, we have

$$\begin{aligned}&\mathbb{E}Q_2(\delta)^*\Sigma_n^{-1}Q_2(\delta) \leq \sigma_0^2\lambda_0^2\|\Sigma_n^{-1}\|_2\sum_{i=1}^{n}\mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}S_{-j}^n(I(\delta)^{\mathcal{D}})(i)\right)^2\\ &= \sigma_0^2\lambda_0^2\|\Sigma_n^{-1}\|_2\sum_{i=1}^{n}\mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}\int_0^{\delta}udX_{t_{n-j-i+1}-u}\right)^2 \qquad (4.72)\\ &= \sigma_0^2\lambda_0^2\|\Sigma_n^{-1}\|_2 n\mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}\int_0^{\delta}udX_{t_{n-j}-u}\right)^2.\end{aligned}$$

For the calculation point of view, we first prove that the stochastic integral satisfies integral by parts as follows, for which the proof can be found in Appendix A.3.1,

$$\mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}\int_0^{\delta}udX_{t_{n-j}-u}\right)^2 = \mathbb{E}\left(\sum_{j=1}^{\infty}e^{(-j+1)\delta\lambda_0}\int_0^{\delta}X_{t_{n-j-1}}-X_{t_{n-j}-u}du\right)^2, \qquad (4.73)$$

and thus

$$
\begin{aligned}
(4.73) &= \mathbb{E}\left(\sum_{j=0}^{\infty} e^{-2j\lambda_0\delta}\Big(\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du\Big)^2\right) \\
&\quad + 2\mathbb{E}\left(\sum_{j=0}^{\infty}\sum_{k>j}^{\infty} e^{-j\lambda_0\delta}e^{-k\lambda_0\delta}\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du\int_0^{\delta}(X_{t_{n-k-1}} - X_{t_{n-k}-u})du\right) \\
&= \mathbb{E}(Q_{2,1}(\delta)) + 2\mathbb{E}(Q_{2,2}(\delta)),
\end{aligned}
\tag{4.74}
$$

where $Q_{2,1}(\delta) := \sum_{j=0}^{\infty} e^{-j\lambda_0\delta}(\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du)^2$, and

$$
Q_{2,2}(\delta) := \sum_{j=0}^{\infty}\sum_{k>j}^{\infty} e^{-j\lambda_0\delta}e^{-k\lambda_0\delta}\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du\int_0^{\delta}(X_{t_{n-k-1}} - X_{t_{n-k}-u})du.
$$

For $\mathbb{E}Q_{2,1}(\delta)$, by Monotone Convergence Theorem (A.1.11), we have

$$
\mathbb{E}(Q_{2,1}(\delta)) = \sum_{j=0}^{\infty} e^{-2j\lambda_0\delta}\mathbb{E}\Big(\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du\Big)^2.
$$

Since obviously, $\mathbb{E}(\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du)^2 < \infty$, by Fubini's theorem (A.1.3),

$$
\begin{aligned}
&\mathbb{E}\Big(\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du\Big)^2 \\
&= \mathbb{E}\left(\int_0^{\delta}\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})(X_{t_{n-j-1}} - X_{t_{n-j}-v})dudv\right) \\
&= \int_0^{\delta}\int_0^{\delta}\mathbb{E}(X_{t_{n-j-1}} - X_{t_{n-j}-u})(X_{t_{n-j-1}} - X_{t_{n-j}-v})dudv \\
&= 2\int_0^{\delta}\int_u^{\delta}\left((\delta-u)^{2h} + (\delta-v)^{2h} - \frac{1}{2}(v-u)^{2h}\right)dvdu \\
&= C(h)\delta^{2h+2}.
\end{aligned}
$$

Thereby

$$\begin{aligned}\mathbb{E}(Q_{2,1}(\delta)) &= \sum_{j=0}^{\infty} e^{-2j\lambda_0\delta}\mathbb{E}(\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du)^2 \\ &\sim O(\delta^{2h+1}).\end{aligned} \tag{4.75}$$

The calculation for $Q_{2,2}(\delta)$ follows the same virtue and we include the calculation in the Appendix A.3.2. From Observation A.3.1, we have

$$\begin{aligned}&\mathbb{E}(Q_{2,2}(\delta)) \\ &= \mathbb{E}\left(\sum_{j=0}^{\infty}\sum_{d=1}^{\infty} e^{-2j\lambda_0\delta}e^{-d\lambda_0\delta}\int_0^{\delta}(X_{t_{n-j-1}} - X_{t_{n-j}-u})du\int_0^{\delta}(X_{t_{n-j-d+1}} - X_{t_{n-j-d+u}})du\right) \\ &= \sum_{j=0}^{\infty}\sum_{d=1}^{\infty} e^{-2j\lambda_0\delta}e^{-d\lambda_0\delta}O((d\delta - \frac{\delta}{2})^{2h-2}\delta^4) \sim O(\delta^3\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}(d\delta)^{2h-2}).\end{aligned} \tag{4.76}$$

Denote $A_\gamma \sim \mathrm{o}(B_\gamma)$, if B_γ is an upper bound of A_γ when $\gamma \to \infty$ or $\gamma \to 0$. For $h < 0.5$,

$$\begin{aligned}(4.76) &\sim O(\delta^3\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}(d\delta)^{2h-2}) \sim O(\delta^{2h+1}\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}d^{2h-2}) \\ &\sim \mathrm{o}(\delta^{2h+1}e^{-\lambda_0\delta}\sum_{d=1}^{\infty} d^{2h-2}) \sim O(\delta^{2h+1}),\end{aligned}$$

and for $h > 0.5$,

$$\begin{aligned}(4.76) &\sim O(\delta^3\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}(d\delta)^{2h-2}) \sim \mathrm{o}(\delta^3\int_{x=0}^{\infty} e^{-x\lambda_0\delta}(x\delta)^{2h-2}dx) \\ &\sim O(\delta^2\int_{y=0}^{\infty} e^{-y\lambda_0}y^{2h-2}dy) \sim \Gamma(2h-1)O(\delta^2).\end{aligned}$$

Therefore by (4.75) and (4.76), we have that for $h > 0.5$,

$$\mathbb{E}(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}\int_0^{\delta} udX_{t_{n-j}-u})^2 \sim O(\delta^2),$$

and for $h < 0.5$.

$$\mathbb{E}(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} \int_0^{\delta} u dX_{t_{n-j}-u})^2 \sim O(\delta^{2h+1}).$$

From Theorem 4.4.6, for $h > 0.5$, $\|\Sigma_n^{-1}\|_2 \sim \delta^{-2h}$, and for $h < 0.5$, $\|\Sigma_n^{-1}\|_2 \sim O(n^{1-2h})\delta^{-2h}$. It is worth noticing that, the theorem is obtained when δ is fixed, and thus we need to consider δ when $\delta \to 0$. Since, from (4.72),

$$\mathbb{E}Q_2(\delta)^*\Sigma_n^{-1}Q_2(\delta) \leq \sigma_0^2\lambda_0^2\|\Sigma_n^{-1}\|_2 n\mathbb{E}\left(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} \int_0^{\delta} u dX_{t_{n-j}-u}\right)^2$$

$$\sim \begin{cases} O(n\delta^{2-2h}), & h > 0.5 \\ O(n^{2-2h}\delta), & h < 0.5 \end{cases}.$$

Recall that (4.71) $= \frac{Q_2(\delta)^*\Sigma_n^{-1}Q_2(\delta)}{n(1-e^{-2\lambda_0\delta})^{-1}}$, and thus (4.71) $\sim O(\delta^{3-2h})$ for $h > 0.5$. For $h < 0.5$, (4.71) $\sim O(n^{1-2h}\delta^2)$ and in order for it to converge to zero, we require that $n^{1-2h}\delta^2 \to 0$, that is

$$n^{\frac{1}{2}-h}\delta \to 0. \tag{4.77}$$

Recall from (4.60) in discrete case, we need the condition $n^{\frac{1}{3}}\delta \to \infty$. Notice that, in order for both (4.77) and (4.70) to satisfy, we have $\frac{1}{2} - h < \frac{1}{3}$, i.e. $h > \frac{1}{6}$. □

The result on $\mathbb{E}\|Q_2(\delta)\|_2$ will be used later, and thus we make a proposition on that, as follows,

Proposition 4.5.1. *$\mathbb{E}\|Q_2(\delta)\|_2^2 \sim O(n\delta^2)$ for $h > 0.5$, and $\mathbb{E}\|Q_2(\delta)\|_2^2 \sim O(n\delta^{1+2h})$ for $h < 0.5$.*

Next, we consider the nominator and prove the following lemma,

Lemma 4.5.6. *We have $\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})} \xrightarrow{P_0} 0$, for $h > 0.5$ with $\delta \to 0$, and $n^{\frac{1}{3}}\delta \to \infty$; for $\frac{1}{3} < h < 0.5$ with $n^{\frac{1}{3}}\delta \to \infty$, $\delta \to 0$, and $\delta n^{1-2h} \to 0$.*

Proof. From Theorem 4.5.5, $\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})}{(1-e^{-2\lambda\delta})^{-1}n} \to 1$ in probability, under some appropriate conditions on δ and n. By Slutsky's theorem (A.1.2), it is equivalent to prove

$$\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta(1-e^{-2\lambda\delta})^{-1}n} \xrightarrow{P_0} 0.$$

By Markov's inequality, we need to prove that

$$\mathbb{E}|\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta(1-e^{-2\lambda\delta})^{-1}n}| \to 0,$$

and by (4.67), we have

$$\begin{aligned}
&\mathbb{E}|\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta(1-e^{-2\lambda\delta})^{-1}n}| \\
&= \mathbb{E}|\frac{(S_{-1}^n(Y(\delta)^{\mathcal{D}})+Q(\delta))^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}}-\lambda_0 I(\delta)^{\mathcal{D}}+\frac{\lambda_0^2}{2}R(\delta)^{\mathcal{D}})}{\delta(1-e^{-2\lambda\delta})^{-1}n}| \\
&= \mathbb{E}|\frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})+Q(\delta)^*\Sigma_n^{-1}(S_0^n(\Delta X(\delta)^{\mathcal{D}}-\lambda_0 I(\delta)^{\mathcal{D}}+\frac{\lambda_0^2}{2}R(\delta)^{\mathcal{D}})}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&+ \frac{S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(-\lambda_0 I(\delta)^{\mathcal{D}}+\frac{\lambda_0^2}{2}R(\delta)^{\mathcal{D}})}{\delta(1-e^{-2\lambda\delta})^{-1}n}| \\
&\leq \mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|+|Q(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|+|Q(\delta)^*\Sigma_n^{-1}\lambda_0 S_0^n(I(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}... \\
&+ \mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}\frac{\lambda^2}{2}S_0^n(R(\delta)^{\mathcal{D}})|+|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}\lambda_0 S_0^n(I(\delta)^{\mathcal{D}})|+|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}\frac{\lambda^2}{2}S_0^n(R(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}
\end{aligned}$$

In the following proof, we repeatedly use the Cauchy's inequality (A.1.7) for random variables. For random variables X and Y, $\mathbb{E}(XY)$ is an inner product [4], and thus by Cauchy's inequality (A.1.7),

$$\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}. \tag{4.78}$$

Since by Jensen's inequality for concave function (A.1.8),

$$\begin{aligned}
&\mathbb{E}|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})| = \mathbb{E}((S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}}))^2)^{\frac{1}{2}} \\
&\leq (\mathbb{E}(S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}}))^2)^{\frac{1}{2}},
\end{aligned}$$

we have, by Lemma 4.5.1,

$$\begin{aligned}
&\frac{\mathbb{E}|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&\leq \frac{1}{\delta(1-e^{-2\lambda\delta})^{-1}}\sqrt{\frac{(\mathbb{E}(S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}}))^2}{n^2}} \to 0.
\end{aligned} \tag{4.79}$$

Since almost surely,

$$R(\delta)^{\mathcal{D}}(i) = \frac{1}{2}\int_0^{\delta} f(u)dX_{t_{n+1-i}-u} \sim O(\delta^2)\Delta X(\delta)^{\mathcal{D}}(i),$$

and

$$I(\delta)^{\mathcal{D}}(i) = \int_0^{\delta} udX_{t_{n+1-i}-u} \sim O(\delta)\Delta X(\delta)^{\mathcal{D}}(i),$$

it is obvious that $\mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$ and $\mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}S_0^n(R(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$ are of smaller order of $\mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$. In addition, from Observation 4.5.2, $\mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(R(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$ is dominated by $\mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$. Thus, we only need to show that the two leading terms $\mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$ and $\mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$ converge to 0 in P_0.
For $\mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$, we have

$$\begin{aligned}
&\mathbb{E}\frac{|Q(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} = \mathbb{E}\frac{|(Q_1(\delta)+Q_2(\delta)+Q_3(\delta))^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&\leq \mathbb{E}\frac{|Q_1(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} + \mathbb{E}\frac{|Q_2(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} + \mathbb{E}\frac{|Q_3(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&= A_1 + A_2 + A_3,
\end{aligned}$$

where

$$A_1 := \mathbb{E}\frac{|Q_1(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n},$$

$$A_2 := \mathbb{E}\frac{|Q_2(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n},$$

and

$$A_3 := \mathbb{E}\frac{|Q_3(\delta)^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}.$$

Recall from (4.69),

$$Q_1(\delta) := (\frac{\lambda_0\delta}{1-e^{-\delta\lambda_0}} - 1)S_{-1}^n(Y(\delta)^{\mathcal{D}}),$$

$$Q_2(\delta) := \sigma_0\lambda_0\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}S_{-j}^n(I(\delta)^{\mathcal{D}}),$$

and

$$Q_3(\delta) := \sigma_0 \frac{\lambda_0^2}{2} \sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0} S_{-j}^n(R(\delta)^{\mathcal{D}}).$$

As A_3 is controlled by A_2, by Observation 4.5.2, we only consider A_1 and A_2. By the same argument in Theorem 4.5.1,

$$A_1 = \mathbb{E}\frac{|Q_1(\delta)^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} \sim O(\delta)\mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{n} \to 0.$$

Recall from Theorem 4.4.6, $\|\Sigma_n^{-1}\|_2 \sim O(\delta^{-2h})$ for $h > 0.5$, and $\|\Sigma_n^{-1}\|_2 \sim O(\delta^{-2h} n^{1-2h})$ for $h < 0.5$. Then for A_2, by Proposition 4.5.1 and (4.78) we have

$$\begin{aligned}
A_2 &= \mathbb{E}\frac{|Q_2(\delta)^* \Sigma_n^{-1} S_0^n(\Delta X(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} \leq \mathbb{E}\frac{\|\Sigma_n^{-\frac{1}{2}}\|_2 \|Q_2(\delta)\|_2 \|Z_0^n\|_2}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&\leq \frac{\|\Sigma_n^{-\frac{1}{2}}\|_2 \sqrt{\mathbb{E}\|Q_2(\delta)\|_2^2} \sqrt{\mathbb{E}\|Z_0^n\|_2^2}}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&\sim \begin{cases} \mathrm{o}(\frac{\delta^{-h} n^{\frac{1}{2}} \delta n^{\frac{1}{2}}}{n}) = \delta^{1-h}, & for \quad h > 0.5 \\ \mathrm{o}(\frac{\delta^{-h} n^{\frac{1}{2}-h} n^{\frac{1}{2}} \delta^{h+\frac{1}{2}} n^{\frac{1}{2}}}{n}) = \delta^{\frac{1}{2}} n^{\frac{1}{2}-h} & for \quad h < 0.5. \end{cases}
\end{aligned}$$

For $h < 0.5$, we need to have $\delta^{\frac{1}{2}} n^{\frac{1}{2}-h} \to 0$, and that is

$$\delta n^{1-2h} \to 0. \tag{4.80}$$

For the other leading term $\mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(I(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n}$, we have, by (4.78) and Jensen's inequality for concave function (A.1.8),

$$\begin{aligned}
&\mathbb{E}\frac{|S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(I(\delta)^{\mathcal{D}})|}{\delta(1-e^{-2\lambda\delta})^{-1}n} \leq \mathbb{E}\frac{\sqrt{S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \sqrt{S_0^n(I(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(I(\delta)^{\mathcal{D}})}}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&\leq \frac{\sqrt{\mathbb{E}(S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}}))} \sqrt{\mathbb{E}(S_0^n(I(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(I(\delta)^{\mathcal{D}}))}}{\delta(1-e^{-2\lambda\delta})^{-1}n} \\
&\sim O(\frac{\sqrt{\mathbb{E}(S_0^n(I(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(I(\delta)^{\mathcal{D}}))}}{\delta\sqrt{(1-e^{-2\lambda\delta})^{-1}n}}) \sim \mathrm{o}(\frac{\sqrt{\|\Sigma_n^{-1}\|_2 \mathbb{E}\|S_0^n(Y(\delta)^{\mathcal{D}})\|_2^2}}{\delta\sqrt{(1-e^{-2\lambda\delta})^{-1}n}}) \\
&\sim \begin{cases} \mathrm{o}(\frac{\delta^{-h} n^{\frac{1}{2}} \delta^{h+1}}{\delta\sqrt{(1-e^{-2\lambda\delta})n}}) = \delta^{\frac{1}{2}}, & for \quad h > 0.5 \\ \mathrm{o}(\frac{\delta^{-h} n^{\frac{1}{2}-h} n^{\frac{1}{2}} \delta^{h+1}}{\delta\sqrt{(1-e^{-2\lambda\delta})n}}) = \delta^{\frac{1}{2}} n^{\frac{1}{2}-h} & for \quad h < 0.5, \end{cases}
\end{aligned}$$

and we need to have for $h < 0.5$,

$$\delta n^{1-2h} \to 0. \tag{4.81}$$

Recall from (4.60) in discrete case, we need the condition $n^{\frac{1}{3}}\delta \to \infty$, and by the above condition, we have $1 - 2h < \frac{1}{3}$, i.e. $h > \frac{1}{3}$. □

We consider the drift parameter λ_0 first, and we prove the following theorem

Theorem 4.5.3. *The approximate MLE for the drifting parameter λ_0 of model 4.29 given by*

$$\tilde{\lambda}_T(\delta) = \frac{-1}{\delta} log \frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(Y^{\mathcal{D}})}{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})},$$

is weakly consistent as $\delta \to 0$ and $\delta n^{\frac{1}{3}} \to \infty$, for $h > 0.5$. It is weakly consistent for $\frac{1}{3} < h < 0.5$, as $\delta n^{\frac{1}{3}} \to \infty$, and $\delta n^{1-2h} \to 0$ with $n \to \infty$ and $\delta \to 0$.

Proof. From Lemma 4.5.5, Lemma 4.5.6, and Slustky's theorem, we have

$$\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta S_{-1}^n(Y(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y(\delta)^{\mathcal{D}})} \xrightarrow{P_0} 0. \tag{4.82}$$

Then, by (4.66), the result is proved. In addition, from Lemma 4.5.5, we need $\delta n^{\frac{1}{2}-h} \to 0$, and from Lemma 4.5.6, we need $\delta n^{1-2h} \to 0$. Since $\delta n^{1-2h} \to 0$ is a stronger condition than $\delta n^{\frac{1}{2}-h} \to 0$, we have the above condition. □

4.5.7 Consistency for the diffusion coefficient of the limiting model

We prove the following theorem

Theorem 4.5.4. *The approximate MLE for the diffusion coefficient σ_0 of model 4.29 given by*

$$\tilde{\sigma}_T^2(\delta) = \frac{\tilde{\lambda}_T^2(\delta)\delta^2}{(1 - e^{-\tilde{\lambda}_T(\delta)\delta})^2} \frac{(S_0^n(Y^{\mathcal{D}}) - S_{-1}^n(Y^{\mathcal{D}})e^{-\tilde{\lambda}_T(\delta)\delta})^* \Sigma_n^{-1} (S_0^n(Y^{\mathcal{D}}) - S_{-1}^n(Y^{\mathcal{D}})e^{-\tilde{\lambda}_T(\delta)\delta})}{n},$$

is weakly consistent as $\delta \to 0$ with $\delta n^{\frac{1}{3}} \to \infty$, for $h > \frac{1}{2}$, and weakly consistent for $\frac{1}{6} < h < \frac{1}{2}$, as $\delta n^{\frac{1}{3}} \to \infty$, $\delta n^{\frac{1}{2}-h} \to 0$ and $n \to \infty$.

Proof. The approximate MLE for σ is as follows,

$$\tilde{\sigma}_T^2(\delta) = \frac{\tilde{\lambda}_T^2(\delta)\delta^2}{(1 - e^{-\tilde{\lambda}_T(\delta)\delta})^2} \frac{(S_0^n(Y^{\mathcal{D}}) - S_{-1}^n(Y^{\mathcal{D}})e^{-\tilde{\lambda}_T(\delta)\delta})^* \Sigma_n^{-1} (S_0^n(Y^{\mathcal{D}}) - S_{-1}^n(Y^{\mathcal{D}})e^{-\tilde{\lambda}_T(\delta)\delta})}{n}.$$

$\frac{\tilde{\lambda}_T^2(\delta)\delta^2}{(1-e^{-\tilde{\lambda}_T(\delta)\delta})^2} \to 1$ as $\delta \to 0$, and we consider the other term,

$$\begin{aligned}
&\frac{(S_0^n(Y^{\mathcal{D}}) - S_{-1}^n(Y^{\mathcal{D}})e^{-\tilde{\lambda}_T(\delta)\delta})^* \Sigma_n^{-1} (S_0^n(Y^{\mathcal{D}}) - S_{-1}^n(Y^{\mathcal{D}})e^{-\tilde{\lambda}_T(\delta)\delta})}{n} \\
&= \frac{\left(S_{-1}^n(Y^{\mathcal{D}})(e^{-\lambda_0\delta} - e^{-\tilde{\lambda}_T(\delta)\delta}) + \sigma_0 S_0^n(\Delta^\lambda X^{\mathcal{D}})\right)^* \Sigma_n^{-1}}{n} \cdots \\
&\left(S_{-1}^n(Y^{\mathcal{D}})(e^{-\lambda_0\delta} - e^{-\tilde{\lambda}_T(\delta)\delta}) + \sigma_0 S_0^n(\Delta^\lambda X^{\mathcal{D}})\right) \\
&= \frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{n} (e^{-\tilde{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})^2 \\
&+ \sigma_0^2 \frac{S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})}{n} \\
&+ 2\frac{S_{-1}^n(Y^{\mathcal{D}}) \Sigma_n^{-1} S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})}{n} (e^{-\tilde{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})\sigma_0 \\
&= A_1 + A_2 + A_3,
\end{aligned} \tag{4.83}$$

where

$$A_1 := \frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{n} (e^{-\tilde{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})^2,$$

$$A_2 := \sigma_0^2 \frac{S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})}{n},$$

and

$$A_3 := 2\frac{S_{-1}^n(Y^{\mathcal{D}}) \Sigma_n^{-1} S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})}{n} (e^{-\tilde{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})\sigma_0.$$

For A_1, we have

$$\begin{aligned}
&\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{n} (e^{-\tilde{\lambda}_T(\delta)\delta} - e^{-\lambda_0\delta})^2 \\
&= \frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{n} \left(\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta^\lambda X^{\mathcal{D}})}{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}\right)^2 \\
&= \frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{n\delta^{-1}} \left(\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta^{\frac{1}{2}} S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}\right)^2.
\end{aligned}$$

From Lemma 4.5.5, we have

$$\frac{S_{-1}^n(Y^{\mathcal{D}})^* \Sigma_n^{-1} S_{-1}^n(Y^{\mathcal{D}})}{n\delta^{-1}} \xrightarrow{P_0} C,$$

where C is a finite number. In addition, from Lemma 4.5.6, we have

$$\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})} \xrightarrow{P_0} 0,$$

and thus

$$\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta^{\frac{1}{2}} S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})} \xrightarrow{P_0} 0.$$

Actually, we can loose the conditions for the case $h < 0.5$ for the above equation. From the proof of (4.80) and (4.81) in Lemma 4.5.6, we require $\delta^{\frac{1}{2}} n^{\frac{1}{2}-h} \to 0$ in order for $\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})} \xrightarrow{P_0} 0$. In our case, the condition changes to

$$\delta n^{\frac{1}{2}-h} \to 0, \tag{4.84}$$

since $\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta^{\frac{1}{2}} S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})}$ is $\delta^{\frac{1}{2}}$ times smaller than $\frac{S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X^{\mathcal{D}})}{\delta S_{-1}^n(Y^{\mathcal{D}})^*\Sigma_n^{-1}S_{-1}^n(Y^{\mathcal{D}})}$. For A_2, we have

$$\begin{aligned}
A2 &= \sigma_0^2 \frac{S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta^\lambda X(\delta)^{\mathcal{D}})}{n} \\
&= \sigma_0^2 \frac{(S_0^n(\Delta X(\delta)^{\mathcal{D}} + \lambda_0 I(\delta)^{\mathcal{D}} + \frac{\lambda_0^2}{2} R(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}(S_0^n(\Delta X(\delta)^{\mathcal{D}} + \lambda_0 I(\delta)^{\mathcal{D}} + \frac{\lambda_0^2}{2} R(\delta)^{\mathcal{D}})}{n} \\
&= \sigma_0^2 (\frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(\Delta X(\delta)^{\mathcal{D}})}{n} + \lambda_0^2 \frac{S_0^n(I(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n} \\
&+ \frac{\lambda_0^4}{4} \frac{S_0^n(R(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(R(\delta)^{\mathcal{D}})}{n} + 2\lambda_0 \frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n} \\
&+ \lambda_0^2 \frac{S_0^n(\Delta X(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(R(\delta)^{\mathcal{D}})}{n} + 2\lambda_0^3 \frac{S_0^n(R(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n}) \\
&\xrightarrow{P_0} \sigma_0,
\end{aligned}$$

by large number theorem, as $n \to \infty$ and $\delta \to 0$. We prove that the remaining terms converge to 0 in P_0, and we only show that the leading term, $\frac{S_0^n(I(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n}$, converges to 0. It is obvious that the cross terms are dominated by the quadratic terms, and $\frac{S_0^n(R(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(R(\delta)^{\mathcal{D}})}{n}$ is of smaller order than $\frac{S_0^n(I(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n}$. By Markov's inequality, we need to prove $\mathbb{E}(\frac{S_0^n(I(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n}) \to 0$, and for which we have

$$\mathbb{E}(\frac{S_0^n(I(\delta)^{\mathcal{D}})^*\Sigma_n^{-1}S_0^n(I(\delta)^{\mathcal{D}})}{n}) \leq \mathbb{E}(\frac{\|\Sigma_n^{-1}\|_2 \|S_0^n(I(\delta)^{\mathcal{D}})\|_2^2}{n})$$

$$\sim \begin{cases} \mathrm{o}(\frac{\delta^{-2h} n \delta^{2h+2}}{n}) = \delta^2, & for \quad h > 0.5 \\ \mathrm{o}(\frac{\delta^{-2h} n^{1-2h} n \delta^{2h+2}}{n}) = \delta^2 n^{1-2h} & for \quad h < 0.5. \end{cases}$$

Combing (4.84), Theorem 4.5.4 is proved. □

4.5.8 Conclusion

In the second project, we derive the approximate MLEs for both the drift coefficient and diffusion coefficients of the discretely observed one-dimensional fractional O-U process, and we proved the weak consistency of the estimators. The method is general enough to include the case where $h < \frac{1}{2}$, and in comparison with other method for discretely observed processes, they only considered the case where $h > \frac{1}{2}$. Some conditions on δ and n are required for the weak consistency. For $h > \frac{1}{2}$, we need $\delta n^{\frac{1}{3}} \to \infty$, this lower bound for δ is from the discrete model where the larger the δ is, the faster the convergence would be. For $h < \frac{1}{2}$, there is an upper bound which is due to the explosive 2 norm of the inverse of the covariance matrix for $h < \frac{1}{2}$, and thus we need an extra condition on δ to make sure the MLEs are consistent.

Chapter 5

Inverse Algorithm

From Chapter 4, we can see that in order to apply the algorithm of calculating MLEs for discretely observed case, we need to construct a piecewise linear driving force given the observations. Since most of the stochastic differential equations admit no closed form of the solutions, we need to construct an algorithm to numerically calculate the piecewise linear driving force. In this section, we study an iterative algorithm solving for the inverse problem, under the condition that the driving force is a piecewise linear path.

The algorithm uses the signature of a path as the feature set instead of the sampling points of the path. It is motivated by the property that the signature on a fixed time segment does not change by adding tree-like paths (p39, Definition 2.28)[41]. Motivated by that property, we create a new path by adding adapted paths, which can be considered as a certain type of tree-like paths, such that the new path goes through the observation points. Since the new path shares the same signature as the original one, the two paths are equivalent in the metric induced by signatures, from which convergence can be achieved.

The chapter is organized as follows. First, we explain the algorithm, and then we apply it to the one-dimensional fractional O-U process as an illustration for the algorithm. Afterwords, we consider differential equations with constant diffusion coefficients and convergence result is obtained for that case. In the end, we combine the results of this chapter and the results from Chapter 4. We conduct numerical analysis on the MLEs for differential equations with no analytical solutions.

5.1 Setting

Suppose that the observations follow the d-dimensional differential equation

$$dY_t = f(Y_t)dt + b(Y_t)dX_t, \quad Y_0 = \bar{c}_0, \tag{5.1}$$

where $f : \mathbb{R}^d \to L(\mathbb{R}, \mathbb{R}^d)$ is Lip(1) and $b : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R}^d)$ is Lip(γ) with $\gamma > p$. The integral is defined in Ito's sense, and the map (5.1) is denoted by $Y_t = I(X_t)$. Suppose that the observations are made on the time points

$$\mathcal{D}_n := \{0, \delta, 2\delta, ..., n\delta\} = \{0, t_1, t_2, ..., t_{n-1}, T\},$$

with $\delta := \frac{T}{n}$ the time between each consecutive observation, and n, the number of observations. The observations on $\mathcal{D}_n$ are denoted by $\bar{c} = \{\bar{c}_{t_i}, t_i \in \mathcal{D}_n\}$ where $\bar{c}_{t_i} = [c^1_{t_i}, c^2_{t_i}, ..., c^d_{t_i}]^*$ is a d-dimensional vector, and $c^j_{t_i}$ refers to the j-th dimension of the the observations at time t_i. In practice, the observations are modelled by continuous stochastic processes, but we often discretize it when applying the model numerically. Thus, it is useful to study the discrete version of model (5.1) given the observations from the (5.1). Let $\{X(n)_t, t \in [0, T]\}$ be the linear interpolation of $\{X_t, t \in [0, T]\}$ on $\mathcal{D}_n$, and we consider the following discrete model

$$dY(n)_t = f(Y(n)_t)dt + b(Y(n)_t)dX(n)_t, \quad Y(n)_0 = \bar{c}_0. \tag{5.2}$$

The aim of the algorithm is to solve for $\{X(n)_t\}$ given observations $\bar{c}$ generated by (5.1), i.e. the algorithm should satisfy the following conditions,

1. $\{Y(n)_t\}$ and $\{X(n)_t\}$ satisfy (5.2)
2. $Y(n)_{t_i} = \bar{c}_{t_i}$, $\forall t_i \in \mathcal{D}_n$
3. $\{X(n)_t\}$ is piecewise linear on $\mathcal{D}_n$.

In order to calculate the process $\{X(n)_t\}$, we need the trajectory of $\{Y(n)_t, t \in [0, T]\}$, whereas we only have discrete points on $\mathcal{D}_n$. An iterated algorithm is designed to construct the path $\{X(n)_t\}$. We denote the outcome for the k^{th} iteration of an general n^{th} piecewise linear path $\{Z(n)_t, t \in [0, T]\}$ by $\{Z^n(k)_t, t \in [0, T]\}$. We set the initial input $\{Y^n(0)_t, t \in [0, T]\}$ as the linear interpolation of $\{Y(n)_t, t \in [0, T]\}$ on $\mathcal{D}_n$. The output $\{X^n(1)_t, t \in [0, T]\}$ might not be piecewise linear, but we force it to be piecewise linear and then apply it to the map $Y^n(1)_t = I(X^n(1)_t)$.

Thus, $\{Y^n(1)_t\}$ might not go through the observations. We want to adjust the path so that it passes through the observations, and at the same time, minimize the measure of the changes of some feature set. Motivated by the property of the signature of a rough path which is that the signature is invariant by adding tree-like paths, the algorithm does the similar adjustment to the path, and we explain the algorithm in the next section.

5.2 The inverse algorithm

Before explaining the algorithm, we make a definition of the adapted path which is motivated by the tree-like path.

Definition 10 (Adapted path)**.** *Given two points X and Y in $\mathbb{R}^d$, the adapted path $\{T_t, t \in [t_i, t_j]\}$ is a $d+1$-dimensional path, such that*

$$T_t = \begin{pmatrix} 0 \\ X + \frac{Y-X}{t_j - t_i}(t - t_i) \end{pmatrix},$$

where $t \in [t_i, t_j]$.

We consider the model (5.2), and in order to exploit the benefit of the property of the signature, we augment the model by making it a (d+1) dimensional differential equation system. Let $\{\tilde{Y}^n(k)_{t'}\}$ be the path generated by adding adapted paths to $\{Y^n(k)_t\}$ such that $\{Y^n(k)_t\}$ goes through the observations, and $\{\tilde{X}^n(k)_{t'}\}$ be the driving force satisfying the following augment the differential equation (5.2),

$$\begin{pmatrix} dY^{(0)}_{t'} \\ d\tilde{Y}^n(k)_{t'} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ f(\tilde{Y}^n(k)_{t'}) & b(\tilde{Y}^n(k)_{t'}) \end{pmatrix} \begin{pmatrix} dX^{(0)}_{t'} \\ d\tilde{X}^n(k)_{t'} \end{pmatrix} \tag{5.3}$$

where $Y^{(0)}_{t'}$ is the time axis, and $t' = \tau(t)$ is an artificial time representing the movement of the path. We denote the above mapping by $I_{aug}(X^{(0)}(n)_{t'}, \tilde{X}^n(k)_{t'})$. Notice that, the adapted paths only add on the observation points, so that the augmented path $\{\tilde{Y}(k)_t\}$ goes from some observation c_{t_i} to $Y(k)_{t_i}$ along the adapted path, and then goes from $Y(k)_{t_i}$ to $Y(k)_{t_{i+1}}$ satisfying the original model (5.2), and then travels from $Y^n(k)_{t_{i+1}}$ to the observation $c_{t_{i+1}}$ along another adapted path. Thus there are two modes of the movements of $\{\tilde{Y}^n(k)_t\}$; one is travelling along adapted paths, and the other one travels according to the original model with an additional dimension, the time dimension.

The augmented differential equation considers the time axis as a dimension of the driving force, and thus we can account for the adapted path to the system by manipulating the movement of $\{X^{(0)}_{t'}\}$. Since the first dimension of the adapted paths is always zero, $dX^{(0)}_{t'}$ is set to be zero for the segment corresponding to the adapted path. An artificial time is constructed to represents the movements of the $d+1$-dimensional path $\{\tilde{Y}^n(k)_{t'}\}$. We give more details on the augmented map by showing explicitly what is the differential equation when $\tilde{Y}^n(k)_{t'}$ is an adapted path and when $\tilde{Y}^n(k)_{t'}$ travels according to the original model.

Adapted path segment We consider the case where the augmented path $\{\tilde{Y}(k)_t\}$ goes from some observation c_{t_i} to $Y(k)_{t_i}$ along the adapted path, and the differential equation is as follows,

$$\begin{pmatrix} d\tilde{Y}^{(0)}_{t'} = dX^{(0)}_{t'} = 0 \\ d\tilde{Y}^n(k)_{t'} = \frac{Y_{t_i} - c_{t_i}}{t'_{i+1} - t'_i} dt' = b(\tilde{Y}^n(k)_{t'}) d\tilde{X}^n(k)_{t'} \end{pmatrix}$$

Travels according to the original model We consider the case where the augmented path $\{\tilde{Y}^n(k)_t\}$ goes from $Y^n(k)_{t_i}$ to $Y^n(k)_{t_{i+1}}$, and the differential equation is as follows,

$$\begin{pmatrix} d\tilde{Y}^{(0)}_{t'} = dX^{(0)}_{t'} = dt' \\ d\tilde{Y}^n(k)_{t'} = f(\tilde{Y}^n(k)_{t'}) dt' + b(\tilde{Y}^n(k)_{t'}) d\tilde{X}^n(k)_{t'}. \end{pmatrix}$$

We can see that $\{\tilde{Y}^n(k)_{t'}\}$ satisfying the original differential equation with an additional time dimension, in a parametrized time t'.
The algorithm can be divided into four steps,

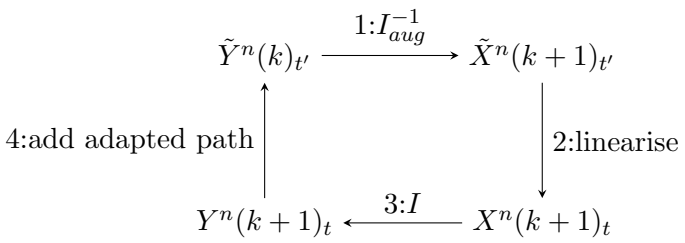
1. $\tilde{X}^n(k+1)_{t'} = I^{-1}_{aug}(\tilde{Y}^n(k)_{t'})$. First, we solve for the inverse of the augmented map to get the driving force.

2. $X^n(k+1)_t = \tilde{X}^n(k+1)_{\tau(t_i)} + \frac{\tilde{X}^n(k+1)_{\tau(t_{i+1})} - \tilde{X}^n(k+1)_{\tau(t_i)}}{\delta}(t - t_i)$, for $t_i \in \mathcal{D}_n$. Then, we linearise the driving force, and set back the artificial time to the original time.

3. $Y^n(k+1)_t = I(X^n(k+1)_t)$. We apply the linearised driving force to the original model.

4. Construct $\{\tilde{Y}^n(k+1)_t\}$ by adding adapted paths to $\{Y^n(k+1)_t\}$ such that $\{\tilde{Y}^n(k+1)_t\}$ passes through the observations.

We illustrate the algorithm by the following diagram,



We make some remarks about the algorithm. For the first step, $\{\tilde{Y}^n(0)_t, t \in [0,T]\}$ is the linear interpolation of the observations,

$$\tilde{Y}^n(0)_t = \bar{c}_{t_i} + \frac{\bar{c}_{t_{i+1}} - \bar{c}_{t_i}}{\delta}(t - t_i),$$

for $t \in [t_i, t_{i+1})$. $\{X^n(1)_t\}$ is solved by solving the inverse of the Ito's map, since $\{\tilde{Y}^n(0)_t\}$ pass through the observations, in which case the the augmented map is identical to Ito's map. In step two, $\{\tilde{X}^n(k)_{t'}\}$ is a piecewise smooth path but might not be piecewise linear, and thus we need to linearise it. In addition, we need to map the artificial time back to the original time. In the third step, we apply Ito's map to $X^n(k)_t$ to obtain $Y^n(k)_t$, and we can see that $Y^n(k)_{t_i} \neq c_{t_i}$ since we do not have the right initial $\{Y^n(0)_t\}$. Therefore, in step 4, we need to adjust $\{Y^n(k)_t\}$ by adding adapted paths, so that the path goes through the observations.

Artificial time and adapted paths We explain more on two important concepts for the algorithm which are adapted paths and the artificial time. Before explaining adapted paths, we need to explain more on the artificial time t'. First, the artificial time relates to the augmented differential equation and adapted paths, and it involves only with step four and step one. In step three, we solve the Ito's map with no artificial time involved. Recall that $t' = \tau(t)$, and we define a homogeneous time point set

$$\mathcal{D}'_n = \{t'_0, t'_1, ..., t'_{3(n-1)-1}\},$$

which includes the ending points of each smooth segment for the augmented path. Suppose we have the right starting observation, and thus, the augmented path passes through the starting point and ending point of the whole path once. It passes through the other $(n-2)$ time points twice. In addition, we need to add

(n-1) observations, and hence the number of artificial time points is

$$|\mathcal{D}'_n| = 2 + 2(n-2) + (n-1) = 3(n-1).$$

As we can see, each time interval $[\tau(t_i), \tau(t_{i+1}))$ for $i \geq 1$, is separated by three time segments, as follows

$$[\tau(t_i), \tau(t_{i+1})) = [\tau(t_i), \tau(t_i) + \delta') \bigcup [\tau(t_i) + \delta', \tau(t_i) + 2\delta') \bigcup [\tau(t_i) + 2\delta', \tau(t_{i+1})),$$

where $\delta' = t'_{i+1} - t'_i = \frac{\tau(T)-\tau(0)}{3(n-1)-1}$. We always consider the behaviour of the paths on time segments of the form, for $i \geq 1$,

$$t' \in [\tau(t_i)+\delta', \tau(t_{i+1})+\delta') = [\tau(t_i)+\delta', \tau(t_i)+2\delta') \bigcup [\tau(t_i)+2\delta', \tau(t_{i+1})) \bigcup [\tau(t_{i+1}), \tau(t_{i+1})+\delta'). \tag{5.4}$$

We define two types of adapted paths on that interval. For $t' \in [\tau(t_i)+\delta', \tau(t_i)+2\delta')$, the adapted path is defined as

$$T_i^-(t'; k+1) = \begin{pmatrix} t_i \\ \bar{c}_{t_i} + \frac{Y^n(k+1)_{t_i} - \bar{c}_{t_i}}{\delta'}(t' - \tau(t_i) - \delta') \end{pmatrix},$$

and for $t' \in [\tau(t_{i+1}), \tau(t_{i+1}) + \delta')$

$$T_i^+(t'; k+1) = \begin{pmatrix} t_i \\ Y^n(k+1)_{t_i} + \frac{\bar{c}_{t_i} - Y^n(k+1)_{t_i}}{\delta'}(t' - \tau(t_i)) \end{pmatrix}.$$

$T_i^+(t'; k+1)$ denotes the kind of tree when the path travels from $Y^n(k+1)_{t_i}$ to the observations, and since we need to adjust the path at t_i so that it goes through the observations, we need to add $T_i^+(t'; k+1)$ for each interval of the form $[\tau(t_i), \tau(t_i)+\delta')$. After the augmented path reaching the observations, it needs to go back to $Y^n(k)_{t_i}$. Hence, we add $T_i^-(t'; k+1)$ on time interval $[\tau(t_i)+\delta', \tau(t_i)+2\delta')$. And $\{\tilde{Y}^n(k+1)_{t'}, t' \in [\tau(t_i)+2\delta', \tau(t_{i+1}))\}$ behaves like $\{Y^n(k+1)_t, t \in [t_i, t_{i+1})\}$. For $i = 0$, we consider the time segment

$$t' \in [\tau(t_0), \tau(t_1) + \delta') = [\tau(t_0), \tau(t_1)) \bigcup [\tau(t_1), \tau(t_1) + \delta'),$$

and we only have $T_1^+(t'; k+1)$ at $\tau(t_1)$. We illustrate the adapted path by the following picture,

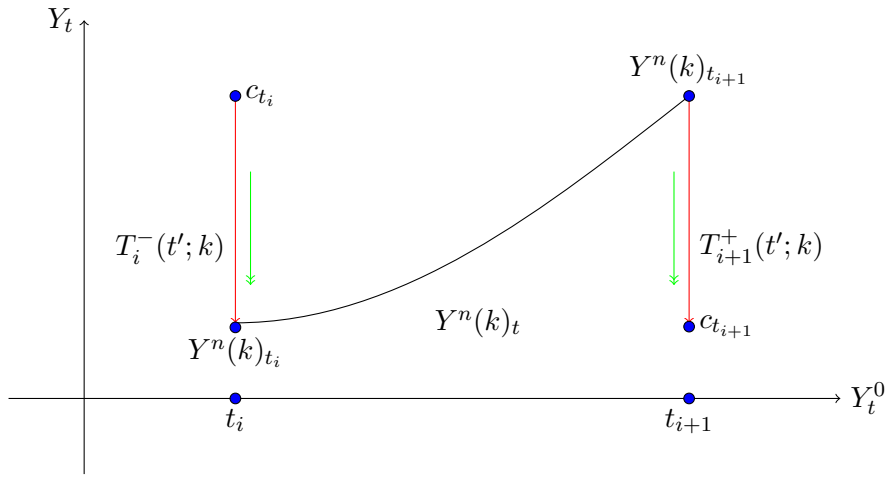


Figure 5.1: 2-dimensional augmented path $\{\tilde{Y}^n(k)_{t'}, t' \in [\tau(t_i)+\delta', \tau(t_{i+1})+\delta']\}$

5.3 Algorithm illustration: 1-dimensional fractional O-U process

In this section, we illustrate the algorithm by applying it to an 1-dimensional fractional O-U process. Let the observations $\{c_{t_i}\}$ be generated by a 1-dimensional fractional O-U process, and the corresponding discrete model be as follows,

$$dY(n)_t = -\lambda Y(n)_t dt + \sigma dX(n)_t, \quad Y(n)_0 = \bar{c}_0, \tag{5.5}$$

where $\lambda,\ \sigma \in \mathbb{R}^+$, and

$$X(n)_t = X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{\delta}(t - t_i),$$

for $t \in [t_i, t_{i+1})$ with $t_i \in \mathcal{D}_n$.

First, we augment the differential equation as follows

$$\begin{pmatrix} dY^{(0)}(n)_{t'} \\ d\tilde{Y}^n(k)_{t'} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\lambda\tilde{Y}^n(k)_{t'} & \sigma \end{pmatrix} \begin{pmatrix} dX^{(0)}(n)_{t'} \\ d\tilde{X}^n(k)_{t'} \end{pmatrix},$$

with $\begin{pmatrix} Y^{(0)}(n)_{t'_0} \\ \tilde{Y}^n(k)_{t'_0} \end{pmatrix} = \begin{pmatrix} 0 \\ c_0 \end{pmatrix}$, and thus

$$\begin{pmatrix} dX^{(0)}(n)_{t'} \\ d\tilde{X}^n(k+1)'_t \end{pmatrix} = \frac{1}{\sigma} \begin{pmatrix} \sigma & 0 \\ \lambda \tilde{Y}^n(k)_{t'} & 1 \end{pmatrix} \begin{pmatrix} dY^{(0)}(n)_{t'} \\ d\tilde{Y}^n(k)_{t'} \end{pmatrix}. \tag{5.6}$$

We consider time interval where $t' \in [\tau(t_i) + \delta', \tau(t_{i+1}) + \delta')$, and we have

$$Y^{(0)}(n)_{t'} = \begin{cases} t_i, \quad t' \in [\tau(t_i) + \delta', \tau(t_i) + 2\delta') \\ t_i + \frac{\delta}{\delta'}(t' - (\tau(t_i) + 2\delta')), \quad t' \in [\tau(t_i) + 2\delta', \tau(t_{i+1})) \\ t_{i+1}, \quad t' \in [\tau(t_{i+1}), \tau(t_{i+1}) + \delta') \end{cases}, \tag{5.7}$$

and

$$\tilde{Y}^n(k)_{t'} = \begin{cases} T_i^-(t'; k) \quad t' \in [\tau(t_i) + \delta', \tau(t_i) + 2\delta') \\ Y^n(k)_{t_i + \frac{\delta}{\delta'}(t' - \tau(t_i) - 2\delta')} \qquad t' \in [\tau(t_i) + 2\delta', \tau(t_{i+1})) \\ T_{i+1}^+(t'; k), \quad t' \in [\tau(t_{i+1}), \tau(t_{i+1}) + \delta') \end{cases}. \tag{5.8}$$

Recall from (4.37),

$$Y(\delta)_t = \sigma \frac{X_{t_{i+1}} - X_{t_i}}{\delta} \int_{t_i}^{t} e^{-\lambda(t-s)} ds + Y(\delta)_{t_i} e^{\lambda(t_i - t)}.$$

We illustrate the algorithm by showing the four steps.

Step One By plugging (5.7) and (5.8) to (5.6), we obtain that

$$\begin{aligned} &\Delta \tilde{X}^n(k+1)_{\tau(t_{i+1})} = X^n(k+1)_{\tau(t_{i+1})+\delta'} - X^n(k+1)_{\tau(t_i)+\delta'} \\ &= \frac{\lambda}{\sigma} \int_{\tau(t_i)+2\delta'}^{\tau(t_{i+1})} \tilde{Y}^n(k)_{t_i + \frac{\delta}{\delta'}(t' - \tau(t_i) - 2\delta')} dY_{t'}^{(0)} + \frac{1}{\sigma}(c_{t_{i+1}} - c_{t_i}) = \frac{\lambda}{\sigma} \int_{t_i}^{t_{i+1}} Y^n(k)_t dt + \frac{1}{\sigma}(c_{t_{i+1}} - c_{t_i}) \\ &= \frac{\lambda}{\sigma} \int_{t=i\delta}^{(i+1)\delta} Y^n(k)_{t_i} e^{-\lambda(t-t_i)} + \sigma \frac{\Delta X^n(k)_{t_{i+1}}}{\lambda\delta}(1 - e^{-\lambda(t-t_i)}) dt + \frac{1}{\sigma}(c_{t_{i+1}} - c_{t_i}) \\ &= \frac{1}{\sigma}(\bar{c}_{t_{i+1}} - \bar{c}_{t_{i+1}} + Y^n(k)_{t_i}(1 - e^{-\lambda\delta})) + \Delta X^n(k)_{t_{i+1}}(1 - \frac{1 - e^{-\lambda\delta}}{\lambda\delta}). \end{aligned} \tag{5.9}$$

Step Two For $t \in [t_i, t_{i+1})$, $X^n(k+1)_t = \tilde{X}^n(k+1)_{\tau(t_{i+1})} + \frac{\Delta \tilde{X}^n(k+1)_{\tau(t_{i+1})}}{\delta}(t-t_i)$, and thus,

$$\Delta X^n(k+1)_{t_{i+1}} = \Delta \tilde{X}^n(k+1)_{\tau(t_{i+1})}.$$

Step Three We plug $\Delta X^n(k+1)_{t_{i+1}}$ to derive $Y^n(k+1)_{t_{i+1}}$,

$$Y^n(k+1)_{t_{i+1}} = \sigma \frac{1-e^{-\lambda\delta}}{\lambda\delta}(\Delta X^n(k+1)_{t_{i+1}}) + Y^n(k+1)_{t_i} e^{-\lambda\delta}. \tag{5.10}$$

Step Four Finally, we add adapted path to $\{Y^n(k+1)_t, t \in [0,T]\}$, and we have

$$\tilde{Y}^n(k+1)_{t'} = \begin{cases} T_i^-(t'; k+1) & t' \in [\tau(t_i)+\delta', \tau(t_i)+2\delta') \\ Y^n(k+1)_{t_i + \frac{\delta}{\delta'}(t'-\tau(t_i)-2\delta')} & t' \in [\tau(t_i)+2\delta', \tau(t_{i+1})) \\ T_{i+1}^+(t'; k+1), & t' \in [\tau(t_{i+1}), \tau(t_{i+1})+\delta') \end{cases}.$$

We repeatedly apply the four steps and stop when the algorithm arrives at a stable point. In the next section, we consider a d-dimensional differential equation with constant diffusion coefficient, and we prove that the algorithm yields a stable point, that is an n-piecewise linear path $\{X_t^n\} := \lim_{k\to\infty}\{X(k)_t\}$, and such path drives the solutions to go through the observations.

5.4 Convergence for processes with constant diffusion coefficient

In this section, we apply the algorithm to the differential equations with constant diffusion coefficient as follows. Suppose we made observations on time point set $\mathcal{D}_n$, and the observations follow the following differential equation

$$dY(n)_t = f(Y(n)_t)dt + \Sigma dX(n)_t, \quad Y(n)_0 = \bar{c}_0, \tag{5.11}$$

where $\{X(n)_t, t = [0,T]\}$ is a d-dimensional n-piecewise linear random process on partition $\mathcal{D}_n$, and the observations of $\{Y(n)_t\}$ are denoted by $\{\bar{c}_t, t \in \mathcal{D}_n\}$. Assume that $f(.) : \mathbb{R}^d \to L(\mathbb{R}, \mathbb{R}^d)$ is a Lipschitz continuous function with Lipschitz coefficient

$$\sup_{X \neq Y} \frac{\|f(X) - f(Y)\|_p}{\|X - Y\|_p} = \lambda_f,$$

where $Y, X \in \mathbb{R}^d$. The l^p metric between $\{Y_t, t = [0,T]\}$ and $\{X_t, t = [0,T]\}$ at time t is defined as

$$\|X_t - Y_t\|_p = (\sum_{j=1}^{d} |Y_t^j - X_t^j|^p)^{\frac{1}{p}},$$

where Y_t^j and X_t^j refer to the j^{th} dimension at time t. We assume that $\Sigma : \mathbb{R}^d \to \mathbb{R}^d$ is invertible with

$$\|\Sigma\|_p = \lambda_\sigma,$$

and

$$\|\Sigma^{-1}\|_p = \lambda_{\sigma^{-1}}.$$

In this section, we prove that the resulting driving force $\{X^n(k)_t, t \in [0,T]\}$ converges in p-variation metric as $k \to \infty$. Before the proof, we give the definition on asymptotic nilpotent matrix.

Definition 11 (Asymptotic nilpotent matrix)**.** *For $A \in \mathcal{C}^{n\times n}$, if $A^k \to \mathbf{0}$ in some matrix norm where $\mathbf{0}$ is a zero matrix as $k \to \infty$, then we say that A is an asymptotic nilpotent matrix.*

Then, we prove a lemma as follows,

Lemma 5.4.1. *Consider the model (5.11). Let $v(\Delta X^n(k+1) - \Delta X^n(k))_p$ denote the vector where*

$$\begin{aligned} &v(\Delta X^n(k+1) - \Delta X^n(k))_p \\ &= [\|\Delta X^n(k+1)_{t_1} - \Delta X^n(k)_{t_1}\|_p, ..., \|\Delta X^n(k+1)_{t_n} - \Delta X^n(k)_{t_n}\|_p]^*, \end{aligned}$$

and we have, $\forall i \in \{1, 2, ..., n\}$,

$$v(\Delta X^n(k+1) - \Delta X^n(k))_p(i) \leq (B(\delta)v(\Delta X^n(k) - \Delta X^n(k-1))_p)\,(i), \quad (5.12)$$

where $B(\delta)$ is an asymptotic nilpotent matrix if $\lambda_\sigma \frac{e^{\lambda_f \delta} - 1 - \lambda_f \lambda_{\sigma^{-1}} \delta}{\lambda_f \lambda_{\sigma^{-1}} \delta} < 1$.

Proof. Based on the algorithm, first we augment the model as follows,

$$\begin{pmatrix} dY^{(0)}(n)_{t'} \\ d\tilde{Y}^n(k)_{t'} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ f(Y^n(k)_t) & \Sigma \end{pmatrix} \begin{pmatrix} dX^{(0)}(n)_{t'} \\ d\tilde{X}^n(k)_{t'} \end{pmatrix}$$

and thus

$$\begin{pmatrix} dX^{(0)}(n)_{t'} \\ d\tilde{X}^n(k)_{t'} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\Sigma^{-1} f(Y^n(k)_t) & \Sigma^{-1} \end{pmatrix} \begin{pmatrix} dY^{(0)}(n)_{t'} \\ d\tilde{Y}^n(k)_{t'} \end{pmatrix}.$$

Recall from (5.4), we always consider the augmented path on the time interval of the following form

$$t' \in [\tau(t_i)+\delta', \tau(t_{i+1})+\delta') = [\tau(t_i)+\delta', \tau(t_i)+2\delta') \bigcup [\tau(t_i)+2\delta', \tau(t_{i+1})) \bigcup [\tau(t_{i+1}), \tau(t_{i+1})+\delta').$$

By the fact that $dY^{(0)}(n)_{t'} = 0$, for $t' \in [\tau(t_i)+\delta', \tau(t_i)+2\delta')$ and $t' \in [\tau(t_{i+1}), \tau(t_{i+1})+\delta')$, and $\tilde{Y}^n(k)_{\tau(t_{i+1})+\delta'} = \bar{c}_{t_{i+1}}$, $\tilde{Y}^n(k)_{\tau(t_i)+\delta'} = \bar{c}_{t_i}$,

$$\begin{aligned} &\tilde{X}^n(k+1)_{\tau(t_{i+1})+\delta'} - \tilde{X}^n(k+1)_{\tau(t_i)+\delta'} \\ &= \Sigma^{-1} \left(- \int_{\tau(t_i)+2\delta'}^{\tau(t_{i+1})} f(\tilde{Y}^n(k)_t) dY_{t'}^{(0)} + \tilde{Y}^n(k)_{\tau(t_{i+1})+\delta'} - \tilde{Y}^n(k)_{\tau(t_i)+\delta'} \right) \\ &= \Sigma^{-1} \left(- \int_{t_i}^{t_{i+1}} f(Y^n(k)_t) dt + \bar{c}_{t_{i+1}} - \bar{c}_{t_i} \right). \end{aligned} \tag{5.13}$$

Because $\tilde{X}^n(k+1)_{\tau(t_{i+1})+\delta'} - \tilde{X}^n(k+1)_{\tau(t_i)+\delta'} = \Delta X^n(k+1)_{t_{i+1}}$, we have the following,

$$\begin{aligned} &\|\Delta X^n(k+1)_{t_{i+1}} - \Delta X^n(k)_{t_{i+1}}\|_p \\ &= \|\Sigma^{-1} \left(\int_{t_i}^{t_{i+1}} f(Y^n(k)_t) - f(Y^n(k-1)_t) dt \right) \|_p \\ &\leq \lambda_{\sigma^{-1}} \| \int_{t_i}^{t_{i+1}} f(Y^n(k)_t) - f(Y^n(k-1)_t) dt \|_p \\ &\leq \lambda_{\sigma^{-1}} \lambda_f \int_{t_i}^{t_{i+1}} \|Y^n(k)_t - Y^n(k-1)_t\|_p dt \\ &= \lambda_{\sigma^{-1}} \lambda_f \int_{t_i}^{t_{i+1}} \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i} ... \end{aligned}$$

$$+\int_{t_i}^{t} f(Y^n(k)_s) - f(Y^n(k-1)_s)ds + \Sigma(\Delta X^n(k)_{t_{i+1}} - \Delta X^n(k-1)_{t_{i+1}})\frac{t-t_i}{\delta}\|_p dt$$

$$\leq \lambda_{\sigma^{-1}}\lambda_f \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p \delta + \lambda_{\sigma^{-1}}\lambda_f \int_{t_i}^{t_{i+1}}\int_{t_i}^{t} \|f(Y^n(k)_s) - f(Y^n(k-1)_s)\|_p ds dt$$

$$+ \frac{\lambda_\sigma \lambda_f \lambda_{\sigma^{-1}}\delta}{2}\|\Delta X^n(k)_{t_{i+1}} - \Delta X^n(k-1)_{t_{i+1}}\|_p.$$

We consider

$$\int_{t_i}^{t_{i+1}}\int_{t_i}^{t} \|f(Y^n(k)_s) - f(Y^n(k-1)_s)\|_p ds dt.$$

Let $I^m(g)_{[t_i,t_{i+1}]}$ denote the m-th level of integral of function $g(t)$ on $[t_i, t_{i+1}]$ as follows,

$$I^m(g)_{[t_i,t_{i+1}]} = \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} g(t_m) dt_m ... dt_1,$$

and

$$\lambda = \lambda_{\sigma^{-1}}\lambda_f. \tag{5.14}$$

We consider $\|f(Y^n(k)_t) - f(Y^n(k-1)_t)\|_p$ as a function of t, and for any $m \in \mathbb{Z}^+$, we have

$$\lambda^{m-1} I^m(\|f(Y^n(k)_t) - f(Y^n(k-1)_t)\|_p)_{[t_i,t_{i+1}]}$$

$$= \lambda^{m-1} \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} \|f(Y^n(k)_{t_m}) - f(Y^n(k-1)_{t_m})\|_p dt_m ... dt_1$$

$$\leq \lambda^m \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} \|Y^n(k)_{t_m} - Y^n(k-1)_{t_m}\|_p dt_m ... dt_1$$

$$= \lambda^m \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i} + \int_{t_i}^{t_m} (f(Y^n(k)_{t_{m+1}}) - f(Y^n(k-1)_{t_{m+1}}))dt_{m+1} +$$

$$+ \Sigma(\frac{\Delta X^n(k)_{t_{i+1}}}{\delta} - \frac{\Delta X^n(k-1)_{t_{i+1}}}{\delta})(t_m - t_i)\|_p dt_m ... dt_1$$

$$\leq \lambda^m \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p dt_m ... dt_1$$

$$
\begin{aligned}
&+\lambda^m \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} \|\Sigma(\frac{\Delta X^n(k)_{t_{i+1}}}{\delta} - \frac{\Delta X^n(k-1)_{t_{i+1}}}{\delta})\|_p (t_m - t_i) dt_m ... dt_1 \\
&+\lambda^m \int_{t_i}^{t_{i+1}} ... \int_{t_i}^{t_{m-1}} \int_{t_i}^{t_m} \|f(Y^n(k)_{t_{m+1}}) - f(Y^n(k-1)_{t_{m+1}})\|_p dt_{m+1} dt_m ... dt_1 \\
&= \lambda^m \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p I^m(1)_{[t_i, t_{i+1}]} \\
&+\lambda_\sigma \lambda^m \|\frac{\Delta X^n(k)_{t_{i+1}}}{\delta} - \frac{\Delta X^n(k-1)_{t_{i+1}}}{\delta}\|_p I^m(t-t_i)_{[t_i, t_{i+1}]} \\
&+\lambda^m I^{m+1}(\|f(Y^n(k)) - f(Y^n(k-1))\|_p)_{[t_i, t_{i+1}]}.
\end{aligned}
$$

Therefore, by iteration, we have

$$
\begin{aligned}
&\lambda \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_1} \|f(Y^n(k)_{t_2}) - f(Y^n(k-1)_{t_2})\|_p dt_2 dt_1 \\
&\leq \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p \sum_{m=2}^{\infty} \lambda^m I^m(1)_{[t_i, t_{i+1}]} \\
&+\lambda_\sigma \frac{\|\Delta X^n(k)_{t_{i+1}} - \Delta X^n(k-1)_{t_{i+1}}\|_p}{\delta} \sum_{m=2}^{\infty} \lambda^m I^m(t-t_i)_{[t_i, t_{i+1}]}.
\end{aligned} \tag{5.15}
$$

Since

$$
I^m(1)_{[t_i, t_{i+1}]} = \frac{(t_{i+1} - t_i)^m}{m!} = \frac{\delta^m}{m!},
$$

and

$$
I^m(t-t_i)_{[t_i, t_{i+1}]} = \frac{(t_{i+1} - t_i)^{m+1}}{(m+1)!} = \frac{\delta^{m+1}}{(m+1)!},
$$

we have

$$
\begin{aligned}
&\lambda \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_1} \|f(Y^n(k)_{t_2}) - f(Y^n(k-1)_{t_2})\|_p dt_2 dt_1 \\
&\leq \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p \sum_{m=2}^{\infty} \lambda^m \frac{\delta^m}{m!} \\
&+\lambda_\sigma \|\Delta X^n(k)_{t_{i+1}} - \Delta X^n(k-1)_{t_{i+1}}\|_p \sum_{m=2}^{\infty} \lambda^m \frac{\delta^m}{(m+1)!}.
\end{aligned} \tag{5.16}
$$

Therefore

$$\begin{aligned}
&\|\Delta X^n(k+1)_{t_{i+1}} - \Delta X^n(k)_{t_{i+1}}\|_p \\
&\leq \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p \sum_{m=1}^{\infty} \lambda^m \frac{\delta^m}{m!} \\
&+ \lambda_\sigma \|\Delta X^n(k)_{t_{i+1}} - \Delta X^n(k-1)_{t_{i+1}}\|_p \sum_{m=1}^{\infty} \lambda^m \frac{\delta^m}{(m+1)!} \\
&= \|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p (e^{\lambda\delta} - 1) + \lambda_\sigma \|\Delta X^n(k)_{t_{i+1}} - \Delta X^n(k-1)_{t_{i+1}}\|_p \frac{e^{\lambda\delta} - 1 - \lambda\delta}{\lambda\delta}.
\end{aligned} \tag{5.17}$$

We would like to find an upper bound for $\|Y^n(k)_{t_i} - Y^n(k-1)_{t_i}\|_p$ controlled by $\{\|\Delta X^n(k)_{t_j} - \Delta X^n(k-1)_{t_j}\|_p, j \leq i\}$. From (5.11), we have for $t \in [t_{i-1}, t_i)$,

$$\begin{aligned}
&Y^n(k)_t - Y^n(k-1)_t \\
&= Y^n(k)_{t_{i-1}} - Y^n(k-1)_{t_{i-1}} + \int_{t_{i-1}}^{t} f(Y^n(k)_s) - f(Y^n(k-1)_s) ds ... \\
&+ \Sigma\left(\frac{\Delta X^n(k)_{t_i}}{\delta} - \frac{\Delta X^n(k-1)_{t_i}}{\delta}\right)(t - t_{i-1}).
\end{aligned} \tag{5.18}$$

Let $U(k)_t = \|Y^n(k)_t - Y^n(k-1)_t\|_p$, and since $\{Y^n(k)_t, t \in [t_{i-1}, t_i)\}$ is continuous,

$$U(k)_t := \|Y^n(k)_t - Y^n(k-1)_t\|_p = \left(\sum_{j=1}^{d} |Y^n(k)_t^j - Y^n(k-1)_t^j|^p\right)^{\frac{1}{p}},$$

is also continuous with respect to t. For $t \in [t_{i-1}, t_i)$,

$$\begin{aligned}
&U(k)_t \\
&\leq U(k)_{t_{i-1}} + \int_{t_{i-1}}^{t} \lambda_f U(k)_s ds + \|\Sigma\left(\frac{\Delta X^n(k+1)_{t_i}}{\delta} - \frac{\Delta X^n(k)_{t_i}}{\delta}\right)\|_p (t - t_{i-1}),
\end{aligned}$$

and by Grönwall's inequality,

$$\begin{aligned}
&U(k)_t \\
&\leq \left(U(k)_{t_{i-1}} + \|\Sigma\left(\frac{\Delta X^n(k)_{t_i}}{\delta} - \frac{\Delta X^n(k-1)_{t_i}}{\delta}\right)\|_p (t - t_i)\right) e^{\lambda_f (t - t_{i-1})}.
\end{aligned}$$

Therefore,

$$\begin{aligned}&U(k)_{t_i}\\&\leq \left(U(k)_{t_{i-1}}+\|\Sigma(\Delta X^n(k)_{t_i}-\Delta X^n(k-1)_{t_i})\|_p\right)e^{\lambda_f\delta},\end{aligned}$$

and by induction,

$$\begin{aligned}&U(k)_{t_i}\\&\leq\sum_{j=0}^{i-1}\|\Sigma(\Delta X^n(k)_{t_{i-j}}-\Delta X^n(k-1)_{t_{i-j}})\|_p e^{(j+1)\lambda_f\delta}\\&\leq\lambda_\sigma\sum_{j=0}^{i-1}\|\Delta X^n(k)_{t_{1+i-j}}-\Delta X^n(k-1)_{t_{1+i-j}}\|_p e^{(j+1)\lambda_f\delta}.\end{aligned}$$

Thus by (5.17), we have

$$\begin{aligned}&\|\Delta X^n(k+1)_{t_{i+1}}-\Delta X^n(k)_{t_{i+1}}\|_p\\&\leq\|Y^n(k)_{t_i}-Y^n(k-1)_{t_i}\|_p(e^{\lambda_f\delta}-1)+\lambda_\sigma\|\Delta X^n(k)_{t_{i+1}}-\Delta X^n(k-1)_{t_{i+1}}\|_p\frac{e^{\lambda_f\delta}-1-\lambda\delta}{\lambda\delta}\\&\leq\lambda_\sigma(e^{\lambda_f\delta}-1)\sum_{j=1}^{i}\|\Delta X^n(k+1)_{t_{i-j}}-\Delta X^n(k)_{t_{i-j}}\|_p e^{j\lambda_f\delta}\\&+\lambda_\sigma\|\Delta X^n(k)_{t_{i+1}}-\Delta X^n(k-1)_{t_{i+1}}\|_p\frac{e^{\lambda_f\delta}-1-\lambda\delta}{\lambda\delta}.\end{aligned}\tag{5.19}$$

By (5.19), we have $\forall i\in\{1,2,...,n\}$,

$$v(\Delta X^n(k+1)-\Delta X^n(k))_p(i)\leq\left(B(\delta)v(\Delta X^n(k)-\Delta X^n(k-1))_p\right)(i),\tag{5.20}$$

where for $i>j$,

$$B(\delta)_{i,j}=\lambda_\sigma e^{(i-j)\lambda_f\delta}(e^{\lambda_f\delta}-1),$$

and

$$B(\delta)_{i,i}=\lambda_\sigma\frac{e^{\lambda_f\delta}-1-\lambda\delta}{\lambda\delta}=\lambda_\sigma\frac{e^{\lambda_f\delta}-1-\lambda_f\lambda_{\sigma^{-1}}\delta}{\lambda_f\lambda_{\sigma^{-1}}\delta}.$$

Let $\lambda_\sigma\frac{e^{\lambda_f\delta}-1-\lambda\delta}{\lambda\delta}<1$, and as n is finite, $B(\delta)_{i,j}<\infty$. Therefore, $B(\delta)$ is an asymptotic nilpotent matrix [6] (Fact 2.4, p3), that is

$$B(\delta)^k\overset{k\to\infty}{\longrightarrow}\mathbf{0}_n,$$

where $\mathbf{0}_n$ is an n by n zero matrix. □

We would like to prove that $\{X^n(k)_t\}$ converges in p-variation metric as $k \to \infty$, and beforehand, we would like to introduce a theorem as follows,

Theorem 5.4.1. *[6] For $A \in \mathcal{C}^{n\times n}$, let $\rho(A) = \max_i \lambda_i$ where λ_i are the eigenvalues of A. If $\rho < 1$, then $A^k \to \mathbf{0}$ in $\|.\|$ where $\mathbf{0}$ is a zero matrix.*

Recall that the p-variation metric is

$$d_p(\mathbf{X}, \mathbf{Y})_{[0,T]} = \max_{i=1,...,\lfloor p \rfloor} \sup_{\mathcal{D}} (\sum_{t_j \in \mathcal{D}} \|\mathbf{X}^i_{t_j,t_{j+1}} - \mathbf{Y}^i_{t_j,t_{j+1}}\|^{\frac{p}{i}})^{\frac{i}{p}},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are two p-rough paths. In our case, $\{X^n(k)_t\}$ is a one-rough path, and then we prove the convergence to some piecewise linear path in 1-variation metric. Since the geometric rough path space is complete, it suffices to prove that the sequence is Cauchy.

Theorem 5.4.2. *For any ϵ, there exists an N, such that for $m,\ k > N$,*

$$d_1(X^n(m), X^n(k))_{[0,T]} < \epsilon.$$

Proof. Given a time point t_i, and let $X^n(k)^j_{t_i}$ be the j^{th} dimension for the vector $X^n(k)_{t_i}$. We take the vector norm for $X^n(k)_{t_i} = [X^n(k)^1_{t_i}, ..., X^n(k)^d_{t_i}]^*$ to be l_q norm where $q \in \mathbb{Z}^+$. By the definition of 1-variation metric, we have

$$d_1(X^n(k), X^n(k-1))_{[0,T]} = \sum_{t^*_j \in \mathcal{D}^*} \|X^n(k)_{t^*_j,t^*_{j+1}} - X^n(k-1)_{t^*_j,t^*_{j+1}}\|_q,$$

where $\mathcal{D}^* = \{t^*_0, t^*_1, ..., t^*_m\}$ is the optimal partition. Since $\{X^n(k)_t - X^n(k-1)_t\}$ is piecewise linear on $\mathcal{D}_n$, $\mathcal{D}^* = \mathcal{D}_n$. Therefore, we have

$$\begin{aligned} d_1(X^n(k), X^n(k-1))_{[0,T]} &= \|v(\Delta X^n(k) - \Delta X^n(k-1))_q\|_1 \\ &\le \|B(\delta)^k v(\Delta X^n(2) - \Delta X^n(1))_q\|_1 \\ &\le \|B(\delta)^k\|_1 \|v(\Delta X^n(2) - \Delta X^n(1))_q\|_1 \end{aligned}$$

where recall that

$$\begin{aligned} &v(\Delta X^n(k) - \Delta X^n(k-1))_q \\ &:= [\|\Delta X^n(k)_{t_1} - \Delta X^n(k-1)_{t_1}\|_q, ..., \|\Delta X^n(k)_{t_n} - \Delta X^n(k-1)_{t_n}\|_q]^*. \end{aligned}$$

Since, by Theorem5.4.1, $\|B(\delta)^l\|_1 \to 0$ as $l \to \infty$, we can find an N, such that $\|B(\delta)^l\|_1 < \frac{\epsilon}{m-k}\|v(\Delta X^n(2) - \Delta X^n(1))_q\|_1^{-1}$ for $l > N$. Thus, for $m,\ k > N$, we have

$$d_1(X^n(m), X^n(k))_{[0,T]} \leq \sum_{i=1}^{m-k} d_1(X^n(k+i), X^n(k))_{[0,T]} < \epsilon.$$

□

Thus, the piecewise linear driving force $\{X^n(k)_t\}$ converges to some piecewise linear path $\{X_t^n\} := \lim_{k\to\infty}\{X^n(k)_t\}$, and we would like to prove such limiting path actually drives the stochastic differential equation to pass through the given observations $\{\bar{c}_{t_i}\}$.

Theorem 5.4.3. *The limiting path $\{X_t^n\} := \lim_{k\to\infty}\{X(k)_t\}$ drives the solution of the differential equation*

$$dY(n)_t = f(Y(n)_t)dt + \Sigma dX_t, \quad Y(n)_0 = \bar{c}_0,$$

to go through the observations $\{\bar{c}_t\}$ at time points $\mathcal{D}_n$.

Proof. Recall from (5.13), that

$$\tilde{X}^n(k+1)_{\tau(t_{i+1})+\delta'} - \tilde{X}^n(k+1)_{\tau(t_i)+\delta'} = \Sigma^{-1}\left(-\int_{t_i}^{t_{i+1}} f(Y^n(k)_t)dt + \bar{c}_{t_{i+1}} - \bar{c}_{t_i}\right),$$

and thus we have

$$\begin{aligned}
\bar{c}_{t_{i+1}} - \bar{c}_{t_i} &= \int_{t_i}^{t_{i+1}} f(Y^n(k)_t)dt + \Sigma(\tilde{X}^n(k+1)_{\tau(t_{i+1})+\delta'} - \tilde{X}^n(k+1)_{\tau(t_i)+\delta'}) \\
&= \int_{t_i}^{t_{i+1}} f(Y^n(k)_t)dt + \Sigma(X^n(k+1)_{t_{i+1}} - X^n(k+1)_{t_i}) \\
&= Y^n(k)_{t_{i+1}} - Y^n(k)_{t_i}.
\end{aligned}$$

Let $k \to \infty$, and we have that the resulting solution goes through the observations.

□

5.5 Numerical Analysis

In this section, we apply the inverse algorithm to construct the approximate MLEs using the method in Chapter 4. In Chapter 4, we apply the method to the fractional O-U process which has an analytical solution. However, for most of the cases, we have to resort to numerical methods to solve the stochastic differential equation. Therefore, it is necessary to study the method by applying it to rough differential equations with no analytical solutions.
We conduct numerical experiments on three cases, the fractional O-U process which has theoretical results and reference for us to compare to. Then we apply it to a two dimensional stochastic model as follows

$$\begin{cases} dY_t^1 = \lambda_{1,1} sin(Y_t^1)dt + \lambda_{1,2} cos(Y_t^2)dt + \sigma_{1,1} dB_t^1 + \sigma_{1,2} dB_t^2 \\ dY_t^2 = \lambda_{2,1} sin(Y_t^1)dt + \lambda_{2,2} cos(Y_t^2)dt + \sigma_{2,1} dB_t^1 + \sigma_{2,2} dB_t^2 \end{cases},$$

where $\{B_t^1, B_t^2\}$ is a two dimensional fractional Brownian motion. Finally, we do experiments on the Cox–Ingersoll–Ross model (CIR) which has non-constant diffusion coefficient,

$$dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}dB_t,$$

where $\{B_t\}$ is a two dimensional fractional Brownian motion.
The section is organized as follows, first, we present the setting for the numerical experiment and then we explain the results of the numerical experiments. The experiments can be categorized into two steps; first, we do experiments on the inverse algorithm and then we do experiments on the calculation of MLEs.

5.5.1 Numerical Setting

We explain some necessary numerical setting for the numerical experiments.

Optimization We calculate MLEs by maximizing the log-likelihood function, and thus we need to find the maximum of the log-likelihood function. We apply a simple optimization algorithm, the line segment method, and the tolerance of error is set to be within 1%. The algorithm is as follows. Consider a function $f(x) : [a, b] \to \mathbb{R}$ and we want to find the maximum of $f(x)$ on an interval $[a, b]$. We consider $-f(x)$, and thus the goal is to find the minimum. The algorithm is as follows,

1. Find $\alpha < \beta < \gamma$ such that $f(\alpha) > f(\beta)$, and $f(\gamma) > f(\beta)$
2. Let $d_1 = \beta + \frac{\gamma-\beta}{2}$. If $f(d_1) > f(\beta)$, $\gamma \leftarrow d_1$; otherwise, $\alpha \leftarrow \beta$, $\beta \leftarrow d_1$, and $f(\beta) \leftarrow f(d_1)$
3. Let $d_2 = \alpha + \frac{\beta-\alpha}{2}$. If $f(d_2) > f(\beta)$, $\alpha \leftarrow d_2$; otherwise, $\gamma \leftarrow \beta$, $\beta \leftarrow d_2$, and $f(\beta) \leftarrow f(d_2)$

where $a \leftarrow b$ means give the value of b to a.

Parameter selection Consider the following differential equation

$$dY_t = a(Y_t; \theta_1)dt + b(Y_t; \theta_2)dX_t, \tag{5.21}$$

where $Y_t \in \mathbb{R}^m$, $a(., \theta_1) : \mathbb{R}^m \to \mathbb{R}^m$, $b(., \theta_2) : \mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^m)$, and $\{X_t\}$ is a m-dimensional fractional Brownian motion. We have the observations on $\{Y_t\}$ at time point set $\mathcal{D}_n = \{0, \delta, 2\delta, ..., n\delta\}$, denoted by $Y^{\mathcal{D}_n}$.
The observations are generated by Euler scheme. Consider $\{Y_t\}$ on $[t_i, t_{i+1})$, and we break the interval $[t_i, t_{i+1})$ into l pieces. Thus, $Y_{t_{i+1}} - Y_{t_i} = \sum_{r=0}^{l-1} (Y_{t_i + \frac{(r+1)\delta}{l}} - Y_{t_i + \frac{r\delta}{l}})$. For $Y_{t_i + \frac{(r+1)\delta}{l}} - Y_{t_i + \frac{r\delta}{l}}$, we have

$$Y_{t_i + \frac{(r+1)\delta}{l}} - Y_{t_i + \frac{r\delta}{l}} = f(\theta; Y_{t_i + \frac{r\delta}{l}}) \frac{\delta}{l} + b(\theta; Y_{t_i + \frac{r\delta}{l}})(X_{t_i + \frac{(r+1)\delta}{l}} - X_{t_i + \frac{r\delta}{l}}).$$

We set $l = 2$ in our numerical experiment. Fractional Brownian motion is generated by fast Fourier transformation [36]. For the inverse algorithm, the number of iteration k is selected such that the l^p norm of the error between two consecutive driving force is within 5% by doing experiment on different k for 20 times for each model. The estimators are the mean value average of 100 Monte-Carlo simulations. The differential equation is solved by the built in function of Matlab, 'ode45' which employs Runge-Kutta method and the absolute error tolerance is set to be within 10^{-7} and relevant error tolerance is 10^{-6}. When solving for $\Delta X^n(k+1)$ given $Y^n(k)$, we need to calculate integrals and we approximate the integral by Riemann-Stieljes approximation. For each interval, we break it down into q pieces, and the integral is calculated by $\Delta X^n(k+1)_{t_i} = \sum_{r=1}^{q} I_{aug}^{-1}(Y^n(k)_{t_{i+(r-1)\delta/q}})(Y^n(k)_{t_{i+r\delta/q}} - Y^n(k)_{t_{i+(r-1)\delta/q}})$. q is decided so that the driving force converges within an acceptable error which is 5% in our case.

Jacobian In order to construct the log-likelihood function, there are two parts we need to calculate for a Gaussian process, which are the Jacobian part and the quadratic part. The quadratic part can be solved by the inverse algorithm, and the construction for the Jacobian for the general model can be referred to [52] (Section 3, p5-p6). We present the result from [52] here. For an m-dimensional process satisfying model (5.21), let $c_{t_i} = \frac{\Delta X_{t_i}}{\delta}$, and we have for $t \in [t_{i-1}, t_i)$,

$$dY_t = a(Y_t;\theta_1)dt + b(Y_t;\theta_2)dX_t = a(Y_t;\theta_1)dt + b(Y_t;\theta_2)c_{t_i}dt.$$

The Jacobian is as follows by equation (15) in [52],

$$\delta^{mn}\prod_{i=1}^{n}|Z(c)_{t_i}|^{-1}, \tag{5.22}$$

where $|Z(c)_{t_i}|$ is the determinate of the matrix of $Z(c)_{t_i}$, and the α^{th}-column of $Z(c)_{t_i}$, $Z^{\alpha}(c)_{t_i}$, admits the following solution,

$$Z^{\alpha}(c)_{t_i} = \int_0^{t_i} A_{s,t}b^{\alpha}(Y(\delta)_s;\theta_2)ds, \tag{5.23}$$

where $\bigtriangledown(a+bc)(Y_t;\theta) := A_t$, and $b^{\alpha}(Y(\delta)_s;\theta_2)$ is the α^{th}-column of the diffusion coefficient $b(Y(\delta)_s;\theta_2)$.

Normalisation of the log-likelihood function Since we have multiple parameters to estimate, we need to scale the log-likelihood function for different parameters, otherwise the parameters of smaller order cannot be calculated accurately. When the log-likelihood function has no theoretical expression, we have to use a rule of thumb for the scaling. Consider the stochastic process (5.1), with parameters $\theta = \{\theta_1, \theta_2\}$, where θ_1 is the set of drift coefficients and θ_2 is the set of diffusion coefficients

$$dY(\delta)_t = f(Y(\delta)_t;\theta_1)dt + b(Y(\delta)_t;\theta_2)c_i dt, \quad Y_0 = y_0.$$

The diffusion coefficients θ_2 dominate the probability measure, since from [55] (Theorem 5), the probability measure $P(Y)$ and $P(\tilde{Y})$ are singular to each other if $\int b_1^2(Y_t;\theta_2)dt \neq \int b_2^2(\tilde{Y}_t;\theta_2)dt$, where $b_1(Y_t;\theta_2)$ and $b_2(\tilde{Y};\theta_2)$ are the diffusion coefficients of two fractional processes $\{Y_t\}$ and $\{\tilde{Y}_t\}$ respectively. In addition, in the

log-likelihood function, $\Delta X^* \Sigma_n^{-1} \Delta X \sim O(n)$ where ΔX is the random vector of the increment of the driving force, almost surely, and for an m-dimensional process, the diffusion coefficient is $O(mn)$. For the Jacobian, by middle value theory, there exists an $s^* \in (t_{i-1}, t_i)$ such that we have

$$Z^\alpha(c)_{t_i} = \int_{t_{i-1}}^{t_i} exp(A_{s,t}) b^\alpha(Y(\delta)_s; \theta_2) ds = exp(A_{s^*,t_i}) b^\alpha(Y(\delta)_{s^*}; \theta_2)\delta \sim O(\delta)\theta_2$$

Thus, $|Z^\alpha(c)| \sim O(\delta^m)\theta_2^m$, and by (5.22), we have the Jacobian is of order θ_2^{mn}. Hence, the Jacobian part involving θ_2 is of order $O(mn)$ in the log-likelihood function, and the diffusion coefficients are expected to be of order $O(mn)$.

We consider the normalisation for drifting coefficients θ_1, and the normalisation process takes an intuition from the O-U process [52]. The log-likelihood function for multidimensional process is the summation of the log-likelihood function of individual dimension, and thus we illustrate the normalisation process by considering a 1-dimensional process. Since θ_1 is of smaller order of θ_2 in the log-likelihood function, we need to eliminate the effect of θ_2, when calculating the MLEs for θ_1. We normalise the log-likelihood function as follows

$$l_T(\theta_1; Y^{\mathcal{D}_n}) = 2l_T(\theta; Y^{\mathcal{D}_{\lfloor \frac{n}{2} \rfloor}}, \theta_2^*) - l_T(\theta; Y^{\mathcal{D}_n}, \theta_2^*),$$

where θ_2^* is the MLE for θ_2, and $Y^{\mathcal{D}_{\lfloor \frac{n}{2} \rfloor}}$ are the observations where the time between each observation is 2δ. Since the log likelihood function can be grouped as follows [52],

$$l_T(\theta; Y^{\mathcal{D}_n}) = \sum_{i=1}^{\infty} \delta^{\alpha_i} l_T^{(\alpha_i)}(\theta_i; Y^{\mathcal{D}_n}),$$

where $l_T^{(\alpha_i)}(\theta_i; Y^{\mathcal{D}_n}) \sim O(1)$, and $\theta = \bigcup \theta_i$, we have

$$\begin{aligned}
&l_T(\theta_1; Y^{\mathcal{D}_n}) = 2l_T(\theta; Y^{\mathcal{D}_{\frac{n}{2}}}, \theta_2^*) - l_T(\theta; Y^{\mathcal{D}_n}, \theta_2^*) \\
&= l_T^{(-1)}(\theta_2^*; Y^{\mathcal{D}_{\frac{n}{2}}}) 2O(\frac{1}{2\delta}) - l_T^{(-1)}(\theta_2^*; Y^{\mathcal{D}_n}) O(\frac{1}{\delta}) \\
&+ l_T^{(\alpha)}(\theta_1; Y^{\mathcal{D}_{\frac{n}{2}}}, \theta_2^*) 2O((2\delta)^\alpha) - l_T^{(\alpha)}(\theta_1; Y^{\mathcal{D}_n}, \theta_2^*) O(\delta^\alpha) \\
&= (l_T^{(-1)}(\theta_2^*; Y^{\mathcal{D}_{\frac{n}{2}}}) - l_T^{(-1)}(\theta_2^*; Y^{\mathcal{D}_n})) O(\frac{1}{\delta}) \\
&+ l_T^{(\alpha)}(\theta_1; Y^{\mathcal{D}_{\frac{n}{2}}}, \theta_2^*) 2O((2\delta)^\alpha) - l_T^{(\alpha)}(\theta_1; Y^{\mathcal{D}_n}, \theta_2^*) O(\delta^\alpha),
\end{aligned}$$

where $\alpha > -1$. Assume that $|l_T^{(-1)}(\theta_2^*; Y^{\mathcal{D}_{\frac{n}{2}}}) - l_T^{(-1)}(\theta_2^*; Y^{\mathcal{D}_n})| \to 0$ as $n \to \infty$, and thus $l_T(\theta_1; Y^{\mathcal{D}_n})$ is a function of order δ^α. The vanishing of the first part depends on the model. We show that actually the intuitive normalisation method works effectively for different models through numerical experiments in the following sections.

5.5.2 Numerical Results

First, we present the numerical results on the inverse algorithm, that is given the observations, we derive the piecewise linear driving force.
Then we calculate the approximate MLEs for the discretely observed models. For a given model, there are three factors for us to adjust, which are T, the observation time, δ, the time interval between observations and n, the number of observations. The behaviour of the estimators depends on their sensibility with respect to the parameters T, δ, and n. From the analytical analysis from previous sections of the log-likelihood function of the fractional O-U process, we make an inference for the general behaviour of the drift parameter and diffusion coefficient. For the first case, we fix T, and change δ. We expect that the volatility σ has better result when δ is smaller, since the estimator in the log-likelihood function is of order $\frac{1}{\delta}$. However, λ converges as $T \to \infty$, and thus the behaviour of the estimator of λ should not change much when T is fixed.
For the second case, we fix n, and change δ. Since, λ converges when $T \to \infty$, λ has better result when δ is bigger. Similar to the first case, σ should be better when δ is smaller.
It is obvious that when we fix δ, both MLEs perform better when n is larger. Thus, we do not do experiment on this case.

5.5.2.1 The inverse algorithm

Based on the previous section, we conclude that the algorithm will yield a convergent piecewise linear driving force for differential equation with constant diffusion coefficients. We would like to present some numerical results on the inverse algorithm to see the convergent behaviour. First, we conduct the numerical experiment on fractional O-U process with $\lambda = 2$, and $\sigma = 1$. We choose $\delta = 0.1$ and $n = 1000$, and we also do experiment on the case where $\delta = 0.01$. Since, from the analysis, the convergence should be dominated by δ, that is a smaller δ will yield a faster convergent rate.

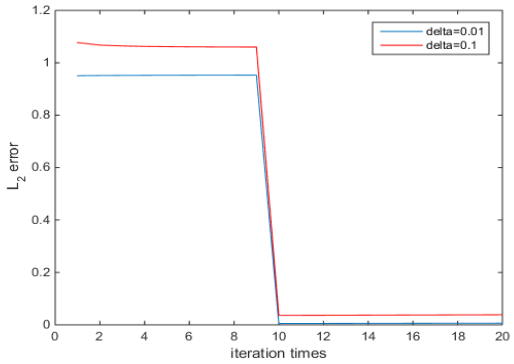


Figure 5.2: The l^2 error between the numerical results and the true paths of the average of 20 paths. $\lambda = 2$, $\sigma = 1$, and $h = 0.6$.

As the plot indicates, there is a sharp turn on the convergence behaviour due to the behaviour of the l^2 norm of $B(\delta)^k$. We illustrate the point by the following numerical results on the l^2 norm of $B(\delta)^k$. Recall from (5.20), $B(\delta)^k$ controls the convergence behaviour of the iterative paths and should converge to $\mathbf{0}$ in matrix norm as $k \to \infty$.

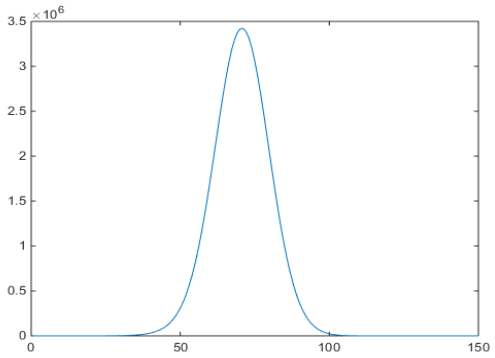


Figure 5.3: The l^2 norm of $B(\delta)^k$ which is $n \times n$ with $n = 200$, $\delta = 0.1$, $\lambda_{\sigma^{-1}} = 3$, $\lambda_f = 1$, and $\lambda_\sigma = 2$. Y-axis is the l^2 norm of $B(\delta)^k$, and X-axis is the number of iteration k.

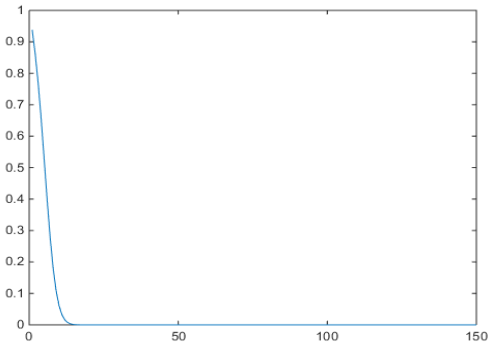


Figure 5.4: The l^2 norm of $B(\delta)^k$ which is $n \times n$ with $n = 200$, $\delta = 0.1$, $\lambda_{\sigma^{-1}} = 3$, $\lambda_f = 1$, and $\lambda_\sigma = 2$. Y-axis is the l^2 norm of $B(\delta)^k$, and X-axis is the number of iteration k.

We proceed to present numerical results on two-dimensional model as follows

$$\begin{cases} dY_t^1 = \lambda_{1,1} sin(Y_t^1)dt + \lambda_{1,2} sin(Y_t^2)dt + \sigma_{1,1} dX_t^1 + \sigma_{1,2} dX_t^2 \\ dY_t^2 = \lambda_{2,1} sin(Y_t^1)dt + \lambda_{2,2} sin(Y_t^2)dt + \sigma_{2,1} dX_t^1 + \sigma_{2,2} dX_t^2 \end{cases}, \quad Y_0 = [0,0]^*,$$

where $\lambda = \begin{bmatrix} 2 & 1.2 \\ 1 & 1.5 \end{bmatrix}$, and $\sigma = \begin{bmatrix} 0.5 & 0.3 \\ 0.2 & 0.4 \end{bmatrix}$. The plot shows a convergent behaviour

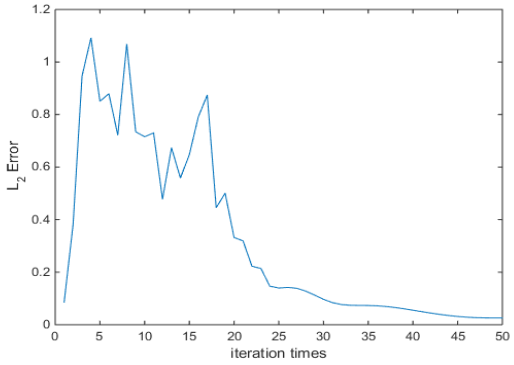


Figure 5.5: The l^2 error between the numerical results and the true paths of the average of 20 paths. $h = 0.6$, $n = 250$, the number of observations, and $\delta = 0.1$.

and a sharp turn around the iteration time of 17, due to the discontinuity of the l^2 norm of the matrix $B(\delta)^k$, and we demonstrate the experimental evidence as follows.

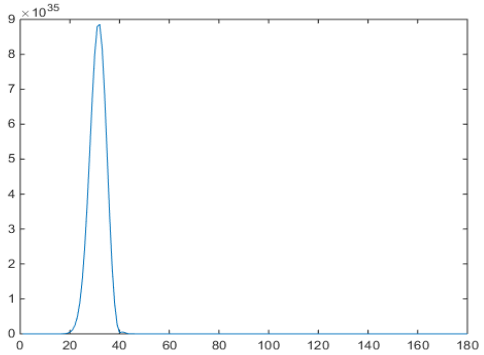


Figure 5.6: The l^2 norm of $B(\delta)^k$ which is $n \times n$ with $n = 200$, $\delta = 0.1$, $\lambda_{\sigma^{-1}} = 3$, $\lambda_f = 1$, and $\lambda_\sigma = 2$. Y-axis is the l^2 norm of $B(\delta)^k$, and X-axis is the number of iteration k.

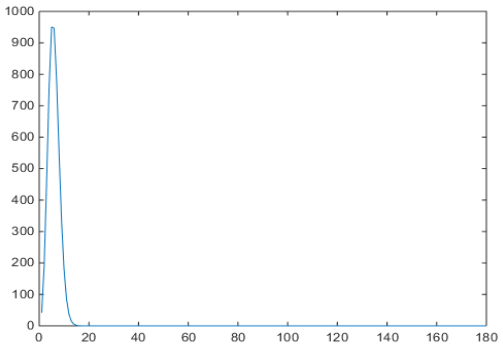


Figure 5.7: The l^2 norm of $B(\delta)^k$ which is $n \times n$ with $n = 200$, $\delta = 0.1$, $\lambda_{\sigma^{-1}} = 3$, $\lambda_f = 1$, and $\lambda_\sigma = 2$. Y-axis is the l^2 norm of $B(\delta)^k$, and X-axis is the number of iteration k.

5.5.2.2 Fractional O-U

Likelihood Comparison Fractional O-U process has analytical solution and thus the likelihood function can be established analytically. We would like to make a comparison between theoretical likelihood and the numerical approximation. The discrepancy lies in two parts. The first one is the quadratic form $S_0^n(\Delta X^{\mathcal{D}})^* \Sigma_n^{-1} S_0^n(\Delta X^{\mathcal{D}})$. The analytical form can be derived by

$$S_0^n(\Delta X^{\mathcal{D}}) = (S_0^n(Y^{\mathcal{D}}) - e^{-\lambda\delta} S_{-1}^n(Y^{\mathcal{D}})) \frac{\lambda\delta}{(1 - e^{-\lambda\delta})\sigma},$$

and we can also obtain the driving force by applying the inverse algorithm. The second error is from Jacobian. The exact Jacobian is $(\frac{\lambda\delta}{(1-e^{-\lambda\delta})\sigma})^n$, but for more general cases, we solve for the Jacobian by solving an ODE numerically. Let $c_{t_i} = \frac{\Delta X_{t_i}}{\delta}$ and $Z_t = \frac{dY_t}{dc_{t_i}}$, we have for $t \in [t_i, t_{i+1})$,

$$\frac{dZ_t}{dt} = \frac{d}{dc_{t_i}}\frac{dY_t}{dt} = \frac{d}{dc_{t_i}}(-\lambda Y_t + \sigma c_{t_i}) = -\lambda Y_t Z_t + \sigma.$$

We did experiments on cases where $h = 0.6$ and $h = 0.4$. Below are the results

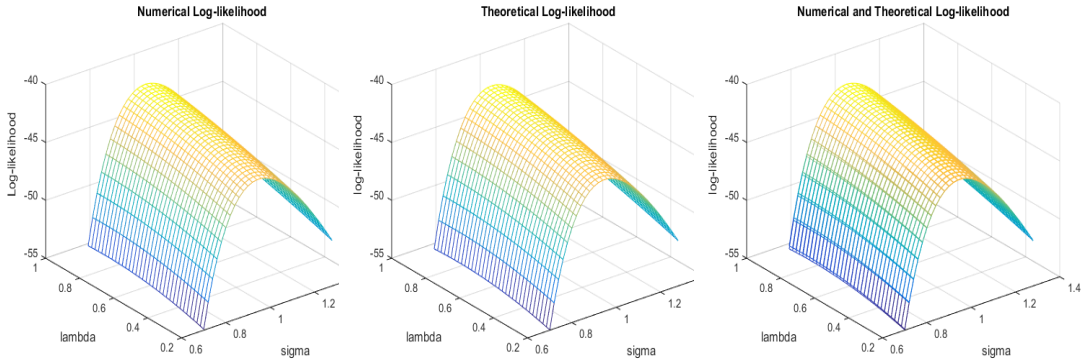


Figure 5.8: The left-hand side picture is the log-likelihood solved numerically for fractional O-U, the middle picture is the exact log-likelihood function for fractional O-U, and the right-hand side picture shows the difference between the two log-likelihood functions. $\lambda = 0.5$, and $\sigma = 1$, and $h = 0.6$. The X-axis is $\lambda = 0.2 : 0.02 : 0.88$, and the Y-axis is $\sigma = 0.7 : 0.02 : 1.28$.

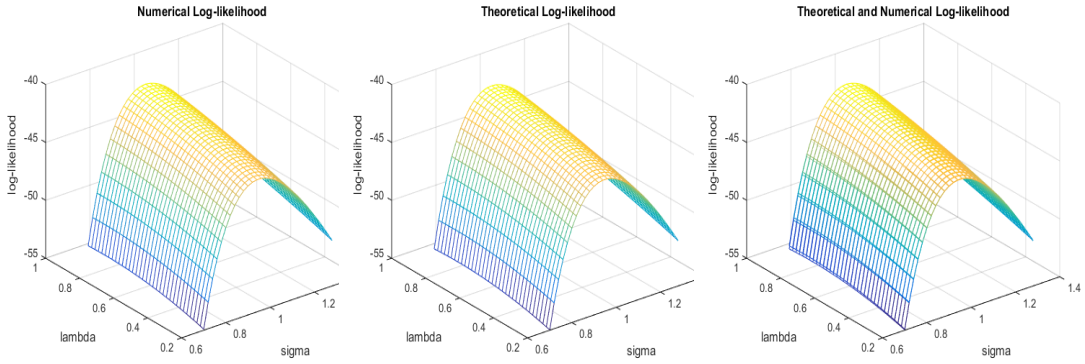


Figure 5.9: The left-hand side picture is the log-likelihood solved numerically for fractional O-U, the middle picture is the exact log-likelihood function for fractional O-U, and the right-hand side picture shows the difference between the two log-likelihood functions. $\lambda = 0.5$, and $\sigma = 1$, and $h = 0.4$. The X-axis is $\lambda = 0.2 : 0.02 : 0.88$, and the Y-axis is $\sigma = 0.7 : 0.02 : 1.28$.

As we can see that the discrepancy is very small for both cases, and the main source of error comes from quadratic form which can be seen from below,

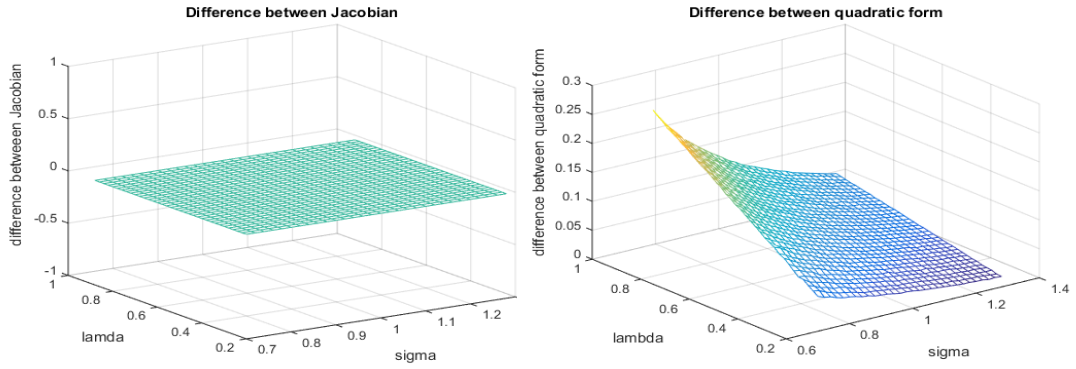


Figure 5.10: The left-hand side picture is difference between the exact and numerical Jacobian , and the right-hand side picture shows the difference between the exact and numerical quadratic forms. $\lambda = 0.5$, and $\sigma = 1$, and $h = 0.6$. The X-axis is $\lambda = 0.2 : 0.02 : 0.88$, and the Y-axis is $\sigma = 0.7 : 0.02 : 1.28$.

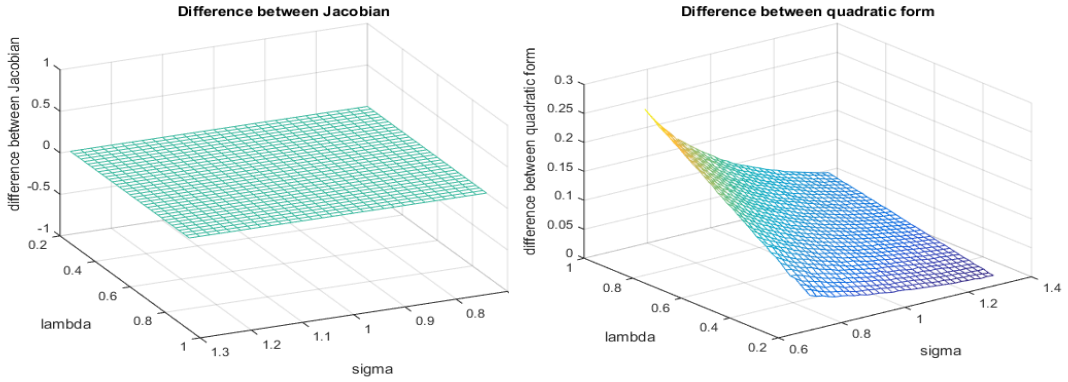


Figure 5.11: The left-hand side picture is difference between the exact and numerical Jacobian , and the right-hand side picture shows the difference between the exact and numerical quadratic forms. $\lambda = 0.5$, and $\sigma = 1$, and $h = 0.4$. The X-axis is $\lambda = 0.2 : 0.02 : 0.88$, and the Y-axis is $\sigma = 0.7 : 0.02 : 1.28$.

MLE We consider the following fractional O-U process

$$dY_t = -0.5Y_t dt + dX_t^h, \qquad Y_0 = 1,$$

where X^h is an h-fractional Brownian motion.

This case serves as an example of comparison, and we construct the setting of the experiment exactly as the one in paper [9], and thus we fix $T = 100$, and $\sigma = 1$ which is known. We present the results by our method first, and below is the results for the first case,

n	$h = 0.4$		$h = 0.6$	
500	0.5118	(0.1090)	0.5227	(0.0797)
1000	0.5084	(0.1086)	0.522	(0.142)

Table 5.1: Mean value average of MLE for λ, and standard deviation in parenthesis for 100 Monte-Carlo simulations. True $\lambda = 0.5$, and $\sigma = 1$ is fixed. $T = 100$ is fixed.

n	$h = 0.4$	$h = 0.6$
500	0.9521(0.1018)	1.0037(0.1414)
1000	0.9738(0.0996)	0.9762(0.1126)

Table 5.2: Mean value average of MLE for λ, and standard deviation in parenthesis for 100 Monte-Carlo simulations. True $\lambda = 1$, and $\sigma = 1$ is fixed. $T = 100$ is fixed.

As we can see, the MLEs for λ shows no significant change when T is fixed, and we can say that δ plays a less significant role in terms of the consistency of λ than T.

For paper[9], the drift parameter λ is constructed using the estimator (4.9) derived from [33],

$$\bar{\lambda}(y)_n = \left(\frac{1}{\sigma^2 H n \Gamma(2H)} \sum_{m=1}^{n} y_m^2 \right)^{\frac{1}{-2H}}.$$

Below is the result from

λ_0	$h = 0.6$
0.5	0.514 (0.166)
1	0.940 (0.238)

Table 5.3: Mean value average for the estimation ofλ, and standard derivation in parenthesis for 500 Monte-Carlo simulations. $\sigma = 1$, $n = 1000$, and $T = 100$.

The results from our method are very similar to the results from above. Next, we perform numerical experiments on the diffusion coefficient. We consider the case where T is fixed with δ changing. Below are the results

n	$h = 0.4$	$h = 0.7$
500	1.0635 (0.0839)	0.8359 (0.0876)
1000	1.0051 (0.1198)	0.9138 (0.1191)

Table 5.4: Mean value average of MLE for σ, and standard derivation in parenthesis for 100 Monte-Carlo simulations. True $\sigma = 1$, and $\lambda = 2$ is fixed. $T = 100$ is fixed.

n	$h = 0.4$	$h = 0.7$
500	2.2030 (0.0999)	2.3576 (0.1495)
1000	2.1791 (0.1041)	2.32 (0.1582)

Table 5.5: Mean value average of MLE for σ, and standard derivation in parenthesis for 100 Monte-Carlo simulations. True $\sigma = 2$, and $\lambda = 2$ is fixed. $T = 100$ is fixed.

The performance is better as δ is smaller for both cases with different Hurst index. The case with smaller Hurst index is slightly better than the one with bigger Hurst index, since the case with bigger Hurst index has long memory, the volatility is not as obvious as the case with smaller Hurst index.
The estimator for σ in paper [9] using the generalised quadratic variation method is as follows,

$$\hat{\sigma}_N = (2 - \frac{V_{N,a}}{\sum_{k,l} a_k a_l |k-l|^{2\hat{h}_N} \Delta^{\hat{h}_N}})^{\frac{1}{2}}.$$

The results are given as follows,

σ_0	$h = 0.7$
1	1.016 (0.282)
2	2.073 (0.564)

Table 5.6: Mean value average of MLE for σ, and standard derivation in parenthesis for 500 Monte-Carlo simulations. $\lambda = 2$, $n = 1000$ and $T = 100$ is fixed.

As we can see that our estimator is more stable comparing to the results from [9] for the diffusion coefficient, and one reason might be that we assume that we know the Hurst index, and the other estimator used an estimated Hurst index in their estimation.
The computation time for our estimators is about a day on the server of the department of statistics, University of Warwick (buster), and since there is no information on the computation time in [9], the efficiency can not be compared. However, it is expected that the estimators from [9] take less time to compute since they do not need to optimise the log-likelihood function. However, their estimators only applies to fractional O-U processes with $h > \frac{1}{2}$, and on the other hand, our method is more general.

5.5.2.3 sine model

$$\begin{cases} dY_t^1 = \lambda_{1,1} sin(Y_t^1)dt + \lambda_{1,2} sin(Y_t^2)dt + \sigma_{1,1} dX_t^1 + \sigma_{1,2} dX_t^2 \\ dY_t^2 = \lambda_{2,1} sin(Y_t^1)dt + \lambda_{2,2} sin(Y_t^2)dt + \sigma_{2,1} dX_t^1 + \sigma_{2,2} dX_t^2 \end{cases}, \quad Y_0 = [0,0]^*,$$

where $\lambda = \begin{bmatrix} 2 & 1.2 \\ 1 & 1.5 \end{bmatrix}$, and $\sigma = \begin{bmatrix} 0.5 & 0.3 \\ 0.2 & 0.4 \end{bmatrix}$. The first experiment is done by fixing T, and changing n. The results for $h = 0.4$ and $h = 0.6$ are shown as below,

n	$\lambda_{1,1}$	$\sigma_{1,1}$
500	1.8866 (0.0705)	0.7120 (0.0207)
1000	1.9080 (0.0306)	0.7002 (0.0681)

Table 5.7: Mean value average of MLE for $\lambda_{1,1}$, and $\sigma_{1,1}$. Standard deviations are in parenthesis for 100 Monte-Carlo simulations. True $\lambda_{1,1} = 2$, and $\sigma_{1,1} = 0.7$. $T = 10$ is fixed. $h = 0.4$

n	$\lambda_{1,1}$	$\sigma_{1,1}$
500	1.8635 (0.0879)	0.7106 (0.0166)
1000	1.9080 (0.0404)	0.7067 (0.0110)

Table 5.8: Mean value average of MLE for $\lambda_{1,1}$, and $\sigma_{1,1}$. Standard deviations are in parenthesis for 100 Monte-Carlo simulations. True $\lambda_{1,1} = 2$, and $\sigma_{1,1} = 0.7$. $T = 10$ is fixed. $h = 0.6$

As we can see, for T fixed, both estimators perform better when δ is smaller. The estimator for the case with smaller Hurst index performs better than the ones with bigger Hurst index, and the volatility shows an obvious improvement when the δ is smaller for $h = 0.4$. Next, we analyse the results from the second experiment where n is fixed with δ changing.

δ	$\lambda_{1,1}$	$\sigma_{1,1}$
0.004	1.8124 (0.1061)	0.6944 (0.0402)
0.04	1.8666 (0.0882)	0.7180 (0.0310)
0.1	1.7945 (0.1045)	0.7502 (0.0403)
0.2	1.8176 (0.1091)	0.7920 (0.0453)

Table 5.9: Mean value average of MLE for λ and σ, and standard derivation in parenthesis for 100 Monte-Carlo simulations. True $\lambda_{1,1} = 2$, and $\sigma_{1,1} = 0.7$. $n = 250$ is fixed. $h = 0.4$

δ	$\lambda_{1,1}$	$\sigma_{1,1}$
0.004	1.8052 (0.1052)	0.6927 (0.0405)
0.04	1.8541 (0.0906)	0.7119 (0.0754)
0.1	1.7915 (0.1030)	0.7445 (0.0387)
0.2	1.8345 (0.1285)	0.8125 (0.0490)

Table 5.10: Mean value average of MLE for λ and σ, and standard derivation in parenthesis for 100 Monte-Carlo simulations. True $\lambda_{1,1} = 2$, and $\sigma_{1,1} = 0.7$. $n = 250$ is fixed. $h = 0.6$

In the second experiment, we fix the number of observations and adjust the time between observations. We can see that, the MLEs for the diffusion coefficient $\sigma_{1,1}$ perform better when δ is smaller. However, the behaviour of $\lambda_{1,1}$ is less obvious. It seems that $\lambda_{1,1}$ achieves the best performance when $\delta = 0.04$, and on the contrary to our expectation, it is not necessary the case that as T increases, the MLEs for $\lambda_{1,1}$ converge to the true value more closely. The explanation is that when δ is too larger, the error due to the lack of the information between each observation also increases. Recall that there is a lower bound for δ in order for the drift coefficient to converge to the true value.

5.5.2.4 CIR model

We consider the Cox–Ingersoll–Ross model which is used to model the interest rate,

$$dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}dX_t^h, r_0 = 0.5,$$

where $\{X_t^h\}$ is an h-fractional Brownian motion and $a = 1.5$, $b = 0.6$, $\sigma = 0.3$. We would like to make a remark on the simulation of the cir paths. Since by [47], the solution of the CIR model is the square root of the fractional O-U process before

hitting zero, that is the process stays at zero after the fractional O-U process hitting zero. Since fractional O-U process admits Riema-Stieljes integral, the corresponding CIR model admits Riemann-Stieljes integral as well and can be simulated using Euler Scheme. Below are the results for the first experiment when $h = 0.6$ and $h = 0.4$ respectively.

n	a	σ
500	1.4083 (0.0433)	0.3016 (0.0129)
1000	1.4063 (0.0455)	0.3012 (0.0087)

Table 5.11: Mean value average of MLE for a, and σ. Standard deviations are in parenthesis for 100 Monte-Carlo simulations. True $a = 1.5$, and $\sigma = 0.3$. $T = 10$ is fixed, and $h = 0.6$.

n	a	σ
500	1.3794 (0.0465)	0.3018(0.0130)
1000	1.3774 (0.0490)	0.3013 (0.0087)

Table 5.12: Mean value average of MLE for a, and σ. Standard deviations are in parenthesis for 100 Monte-Carlo simulations. True $a = 1.5$, and $\sigma = 0.3$. $T = 10$ is fixed, and $h = 0.4$.

For the first experiment, we fix T, and adjust the number of observations. The results indicate that the case where $h = 0.6$ performs better than the case where $h = 0.4$. For both cases, MLEs for a behaves similar when T is fixed, and for σ, the results are better when δ is smaller. Next, we consider the second experiment, where we fix n, and change δ. Below are the results,

δ	a	σ
0.004	1.3986 (0.046)	0.3002 (0.0168)
0.04	1.3974 (0.0451)	0.3047 (0.0168)
0.2	1.3724 (0.0427)	0.3196 (0.0166)

Table 5.13: Mean value average of MLE for a, and σ. Standard deviations are in parenthesis for 100 Monte-Carlo simulations. True $a = 1.5$, and $\sigma = 0.3$. $n = 250$ is fixed, and $h = 0.4$.

δ	a	σ
0.004	1.4078 (0.0446)	0.3001 (0.0171)
0.04	1.4057 (0.0418)	0.3042 (0.0172)
0.2	1.4302 (0.0168)	0.3160 (0.0169)

Table 5.14: Mean value average of MLE for a, and σ. Standard deviations are in parenthesis for 100 Monte-Carlo simulations. True $a = 1.5$, and $\sigma = 0.3$. $n = 250$ is fixed, and $h = 0.6$.

In the second experiment, we fix the number of observations and adjust the time between observations. We can see that, the MLEs for the diffusion coefficients perform better when δ is smaller. However, the behaviour of a is less obvious, and on the contrary to our expectation, it is not necessary the case that as T increases, the MLEs for a converge to the true value more closely. The reason for the observation is that when δ is too larger, the error due to the lack of the information between each observation also increases. This will make the estimation less accurate. Recall that there is a lower bound for δ in order for the drift coefficient to converge to the true value.

5.5.3 Error of the numerical experiment

We make some discussion on the sources of error in the numerical experiment for the calculation of MLEs. First source is from the optimization algorithm, since the algorithm might converge around the monotonic segment of the function instead of the minimum. The second possible source is from the calculation of the driving force. The driving force cannot be calculated accurately, which cause the error in the estimation.

5.6 Conclusion

In this chapter, numerical scheme of calculating the driving force is given, and in addition to the previous chapter, numerical experiments on the construction of approximate MLEs given discrete observations are done. Analytical results for the convergence of the path for differential equation with constant diffusion coefficients is derived, but the convergence result for more general cases can be discussed. In addition, for the algorithm, different ways of constructing the starting point are possible, and the sensitivity of the starting point to the algorithm can be explored

in the future.

Appendix A

Appendix

A.1 Theorem

Theorem A.1.1 (Taylor's expansion theory). *[58](Theorem 5.15, p110) Let f be an $(n+1)$ times differentiable function on an open interval which contains points a and x. Then*

$$f(x) = f(a) + f(a)'(x-a) + \frac{f^2(a)}{2!}(x-a)^2 + ...\frac{f^n(a)}{n!}(x-a)^n + R_n(x),$$

where

$$R_n(x) = \frac{f^{n+1}(c)}{(n+1)!}(x-a)^{n+1}$$

for some number c between a and x.

Theorem A.1.2 (Slutsky's Theorem). *[65](Theorem 2.8, p11) If X_n converges in distribution to a random element X, and Y_n converges in probability to a constant c, then*

1. *$X_n + Y_n \to X + c$ in distribution*
2. *$X_n Y_n \to Xc$ in distribution*
3. *$X_n/Y_n \to X/c$ in distribution provided $c \neq 0$.*

Theorem A.1.3 (Fubini's Theorem). *[3](Chapter 5.2, p152) Let A and B are σ finite measure, and if any of the following is true*

$$\int_A \int_B |f(x,y)|dxdy < \infty, \qquad \int_B \int_A |f(x,y)| < \infty,$$

then

$$\int_A \int_B f(x,y)dxdy = \int_B \int_A f(x,y)dydx = \int_{A\times B} f(x,y)d(x\times y).$$

Theorem A.1.4 (Grownwall'sinequality)**.** *[43] Let $g(t)$ and $f(t)$ be non-negative continuous functions on $[0,T]$, and α be a positive number. If*

$$f(t) \leq g(t) + \alpha \int_0^t f(s)ds,$$

then

$$f(t) \leq g(t) + \alpha \int_0^t g(s)e^{\alpha(t-s)}ds.$$

Theorem A.1.5 (Dominated Convergence Theorem)**.** *[67] If $X_n \to X$, and $|X_n| < Y$ for all n almost surely where Y is integrable, then*

$$\lim_{n\to\infty} E(X_n) = E(X).$$

Theorem A.1.6 (Isserli's Theorem)**.** *[45] If $(X_1, X_2, ..., X_n)$ is a zero mean multivariate normal vector, then*

$$E(X_1X_2...X_{2n}) = \sum\prod E(X_iX_j),$$

$$E(X_1X_2...X_{2n+1}) = 0,$$

where $\sum\prod$ means summing over all the distinct ways of partitioning $(X_1, X_2, ..., X_n)$ into product of pairs X_iX_j.

Theorem A.1.7 (Cauchy's Inequality)**.** *[62](Proposition 1.2.4, p3) Let $< ., . >$ be the inner product for some vector space V, and for $x, y \in V$, we have*

$$< x, y >^2 \leq < x, x >< y, y > .$$

Theorem A.1.8 (Jensen's Inequality)**.** *If X is a random variable and ϕ is a concave function, then*

$$\phi(\mathbb{E}(X)) \geq \mathbb{E}(\phi(X)).$$

.

Proof. Suppose f is differentiable. The function f is concave if

$$f(x) \leq f(y) + (x - y)f'(y).$$

Let $x = X$ and $y = \mathbb{E}(X)$, and we have

$$f(X) \leq f(E[X]) + (X - \mathbb{E}(X))f'(\mathbb{E}(X)).$$

By taking expectation on both sides, we have

$$\mathbb{E}(f(X)) \leq f(\mathbb{E}(X)) + \mathbb{E}(X - \mathbb{E}(X))f'(\mathbb{E}(X)) = f(\mathbb{E}(X)).$$

□

Theorem A.1.9 (Continuous mapping theorem)**.** *[3](Theorem 9.4.2, p305) Let $g : \mathbb{R}^k \to \mathbb{R}^m$ be continuous at every point of a set C such that $P(X \in C) = 1$. If $X_n \to X$ in probability, then $g(X_n) \to g(X)$ in probability.*

Theorem A.1.10 (Markov's inequality)**.** *[3](Chapter 3, p83) Let X be a random variable, and for some $p > 0$ and $M >$), it follows,*

$$P(|X| > M) \leq \frac{E|X|^p}{M^p}$$

Theorem A.1.11 (Monotone Convergence Theorem)**.** *[67](Chapter 5) Suppose that $(X_n : n \geq 0)$ is a sequence of non-negative monotonic random variables with $X_\infty = \lim_{n\to\infty} X_n$ and then $\mathbb{E}(X_n) \to \mathbb{E}(X_\infty)$.*

A.2 Calculations for section 4.5.3

A.2.1 Calculation for inequality (4.52)

$$P_0(|\frac{N_n}{n}| > \epsilon) = P_0(|\frac{< \sum_{k=0}^{\infty} e^{-k\lambda_0\delta} Z_{-k-1}^n, Z_0^n >}{n}| > \epsilon)$$
$$\leq \frac{1}{n^2\epsilon^2}\mathbb{E}\left(< \sum_{k=0}^{\infty} e^{-k\lambda_0\delta} Z_{-k-1}^n, Z_0^n >\right)^2$$

$$
\begin{aligned}
&= \frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=0}^{\infty} e^{-k\lambda_0\delta} < Z_{-k-1}^n, Z_0^n >\right)^2 \\
&= \frac{e^{\lambda_0\delta}}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=1}^{\infty} e^{-k\lambda_0\delta} G_k(n)\right)^2 \\
&= \frac{e^{\lambda_0\delta}}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=1}^{\infty} e^{-2k\lambda_0\delta} G_k(n)^2 + 2\sum_{k=1}^{\infty}\sum_{l>k}^{\infty} e^{-(k+l)\lambda_0\delta} G_k(n)G_l(n)\right) \\
&\leq \frac{e^{\lambda_0\delta}}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=1}^{\infty} e^{-2k\lambda_0\delta} G_k(n)^2 + \sum_{k=1}^{\infty}\sum_{l>k}^{\infty} e^{-(k+l)\lambda_0\delta}(G_k(n)^2 + G_l(n)^2)\right) \qquad \text{(a)} \\
&= \frac{e^{\lambda_0\delta}}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=1}^{\infty} e^{-2k\lambda_0\delta} G_k(n)^2 + \sum_{k=1}^{\infty} e^{-2k\lambda_0\delta} G_k(n)^2\frac{e^{-\lambda_0\delta}-1}{1-e^{-\lambda_0\delta}} + \sum_{k=1}^{\infty} e^{-k\lambda_0\delta} G_k(n)^2\frac{e^{-\lambda_0\delta}}{1-e^{-\lambda_0\delta}}\right) \\
&\sim O(\frac{e^{-\lambda_0\delta}}{1-e^{-\lambda_0\delta}}\frac{1}{n^2\epsilon^2}\mathbb{E}\left(\sum_{k=1}^{\infty} e^{-k\lambda_0\delta} G_k(n)^2\right)) \qquad \text{(b)} \\
&\sim O(\frac{e^{-\lambda_0\delta}}{1-e^{-\lambda_0\delta}}\frac{1}{n^2\epsilon^2}\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}\mathbb{E}G_k(n)^2).
\end{aligned}
$$

We explain step (a) in more detail,

$$
\begin{aligned}
&\sum_{k=1}^{\infty}\sum_{l>k}^{\infty} e^{-(k+l)\lambda_0\delta}(G_k(n)^2 + G_l(n)^2) = \sum_{k=1}^{\infty}\sum_{l>k}^{\infty} e^{-(k+l)\lambda_0\delta} G_k(n)^2 + \sum_{k=1}^{\infty}\sum_{l>k}^{\infty} e^{-(k+l)\lambda_0\delta} G_l(n)^2 \\
&= \sum_{k=1}^{\infty} e^{-k\lambda_0\delta} G_k(n)^2 \sum_{k=l+1}^{\infty} e^{-l\lambda_0\delta} + \sum_{l=1}^{\infty} e^{-l\lambda_0\delta} G_l(n)^2 \sum_{k=1}^{l-1} e^{-k\lambda_0\delta} \\
&= \sum_{k=1}^{\infty} e^{-k\lambda_0\delta} G_k(n)^2 \frac{e^{-(k+1)\lambda_0\delta}}{1-e^{-\lambda_0\delta}} + \sum_{l=1}^{\infty} e^{-l\lambda_0\delta} G_l(n)^2 \frac{e^{-\lambda_0\delta}(1-e^{-(l-1)\lambda_0\delta})}{1-e^{-\lambda_0\delta}} \\
&= \sum_{k=1}^{\infty} e^{-2k\lambda_0\delta} G_k(n)^2 \frac{e^{-\lambda_0\delta}-1}{1-e^{-\lambda_0\delta}} + \sum_{k=1}^{\infty} e^{-k\lambda_0\delta} G_k(n)^2 \frac{e^{-\lambda_0\delta}}{1-e^{-\lambda_0\delta}}.
\end{aligned}
$$

In step (b), the reason we can swap the summation and the limit operation is due to the Monotone Convergence Theorem (A.1.11).
We expand $\mathbb{E}(G_k(n))^2$ as follows, using Isserli's Theorem (A.1.6),

$$
\begin{aligned}
\mathbb{E}(G_k(n))^2 &= \mathbb{E}(S_{-d}^n(\Delta X(\delta)^{\mathcal{D}})^* \Sigma_n^{-1} S_{-d-k}^n(\Delta X(\delta)^{\mathcal{D}}))^2 \\
&= \mathbb{E}(Z_{-d}^{n*} Z_{-d-k}^n)^2 = \mathbb{E}(\sum_{i=1}^{n} Z_{-d}^n(i) Z_{-d-k}^n(i))^2
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Z^n_{-d}(i)Z^n_{-d-k}(i)Z^n_{-d}(j)Z^n_{-d-k}(j)\right)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d}(j))\mathbb{E}(Z^n_{-d-k}(i)Z^n_{-d-k}(j))\\
&+ \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(i))\mathbb{E}(Z^n_{-d}(j)Z^n_{-d-k}(j))\\
&+ \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(Z^n_{-d}(i)Z^n_{-d-k}(j))\mathbb{E}(Z^n_{-d}(j)Z^n_{-d-k}(i)).
\end{aligned}
$$

A.2.2 Properties on the matrix V_k

We define a matrix V_k as follows,

$$V_k := \mathbb{E}(S^n_{-d}(\Delta X^{\mathcal{D}})S^n_{-d-k}(\Delta X^{\mathcal{D}})^*) = \mathbb{E}(S^n_0(\Delta X^{\mathcal{D}})S^n_{-k}(\Delta X^{\mathcal{D}})^*),$$

and thus,

$$V_k(i,j) = \mathbb{E}(\Delta X_{t_{n-i}}\Delta X_{t_{n-k-j}}) = \rho_{|k+j-i|}. \tag{A.1}$$

Notice that $V_k = V^*_{-k}$. Next, we introduce some notations.

1. $d(A)$ denotes the set of diagonal elements of matrix A
2. $r_k(A)$ denotes the k^{th} row of matrix A
3. $c_k(A)$ denotes the k^{th} column of matrix A.

We make an observation on matrix $V_d\Sigma_n^{-1}$. Recall that Σ_n is the covariance matrix of $S^n_0(\Delta X(\delta)^{\mathcal{D}})$,

$$\Sigma_n = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & \rho_0 \end{bmatrix},$$

where $\rho_i = \mathbb{E}(S^n_0(\Delta X(\delta)^{\mathcal{D}})(j)S^n_0(\Delta X(\delta)^{\mathcal{D}})(j \pm i))$, and Σ_n^{-1} is its inverse, denoted as follows,

$$\Sigma_n^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{bmatrix}.$$

Observation A.2.1. *$d(V_k\Sigma_n^{-1})$ has k number of non-zero elements for $0 < k \leq n$.*

Proof. For $d \geq 1$, by (A.1), we have $V_k(d+k, j) = \rho_{|j-d|} = \Sigma_n(d, j)$, and thus, $r_{d+k}(V_k) = r_d(\Sigma_n)$. Therefore,

$$\begin{aligned} V_k\Sigma_n^{-1}(d+k, d+k) &=< r_{d+k}(V_k), c_{d+k}(\Sigma_n^{-1}) > \\ &=< r_d(\Sigma_n), c_{d+k}(\Sigma_n^{-1}) > \\ &= 0, \end{aligned}$$

that is for the elements $V_k\Sigma_n^{-1}(j, j)$ where $j > k$ is zero. □

We would like to prove that $V_d\Sigma_n^{-1}(i, j) < \infty$.

Theorem A.2.1. *$V_d\Sigma_n^{-1}(i, j) < \infty$ for $d \in \mathbb{Z}$, and $1 < i, j < n$.*

Proof. We prove for the case where $d > 0$, and the case where $d < 0$ follows immediately, since $V_d = V_{-d}^*$. The proof can be broken into two scenarios. For $d < n$ and $i \geq (d+1)$, the i^{th} row of V_d is $r_i(V_d) = r_{i-d}(\Sigma_n)$. Therefore,

$$< r_i(V_d), c_j(\Sigma_n^{-1}) >=< r_{i-d}(\Sigma_n), c_j(\Sigma_n^{-1}) >= \begin{cases} 1, & j = i - d \\ 0, & otherwise. \end{cases}$$

For other scenarios which are $d \leq n$ and $i \leq (d+1)$, and $d > n$, $r_i(V_d) = [\rho_{d-i+1}, \rho_{d-i+2}, ..., \rho_{d-i+n}]$. We consider the following series

$$V_d\Sigma_n^{-1}(i, j) = \sum_{k=1}^{n} \rho_{d-i+k} a_{k,j} = \sum_{k=1}^{n} \frac{\rho_{d-i+k}}{r_1(\Sigma_n)(k)} r_1(\Sigma_n)(k) a_{k,j} = \sum_{k=1}^{n} b_{d,i}(k) l_j(k),$$

where $b_{d,i}(k) := \frac{\rho_{d-i+k}}{r_1(\Sigma_n)(k)}$, and $l_j(k) := r_1(\Sigma_n)(k) a_{k,j}$. Recall that

$$\begin{aligned} \rho_{|i-j|} &= \mathbb{E}(\Delta B_{t_i}^h \Delta B_{t_{n-j}}^h) \\ &= \frac{\delta^{2h}}{2}\{(|i-j|+1)^{2h} + (|i-j|-1)^{2h} - 2|i-j|^{2h}\}, \end{aligned}$$

and from (3.2), when $m >> 1$,

$$\rho_m \sim \frac{\delta^{2h}}{2} O(m^{2h-2}).$$

Therefore, for $d-i << n-1$, by Taylor's expansion at $n-1$, we have

$$b_{d,i}(n) = \frac{\rho_{d-i+n}}{\rho_{n-1}} \sim O(\frac{(d-i+n)^{2h-2}}{(n-1)^{2h-2}}) \sim O(1+n^{-1}). \tag{A.2}$$

$\sum_{k=1}^{n} l_j(k)$ is convergent, since if $j=1$, $\sum_{k=1}^{n} l_j(k) = 1$, otherwise $\sum_{k=1}^{n} l_j(k) = 0$. By applying summation by parts to $V_d\Sigma_n^{-1}(i,j)$, we have

$$V_d\Sigma_n^{-1}(i,j) = \sum_{k=1}^{n} b_{d,i}(k)l_j(k) = \sum_{k=1}^{n-1}(b_{d,i}(k) - b_{d,i}(k+1))L_j(k) + b_{d,i}(n)L_j(n), \tag{A.3}$$

where $L_j(k) = \sum_{r=1}^{k} l_j(r)$. Since from (A.2), $b_{d,i}(k) - b_{d,i}(k+1) \sim O(k^{-2})$, $L_j(k) < \infty$, and $b_{d,i}(n)L_j(n) \to L_j(n) < \infty$, the above series converges.
For $d-i \sim O(n-1)$, $b_{d,i}(n) \sim C(n)^{2h-2}$ with $C(n)^{2h-2}$ a constant depending on n. Thus, $b_{d,i}(k) - b_{d,i}(k+1) \sim O(k^{2h-3}) \sim \mathrm{o}(k^{-1})$ for $h < 1$, and by the same same argument in the earlier case, (A.3) converges. It is obvious that for for $d-i >> n-1$, (A.3) converges. Since $b_{d,i}(n) \sim O(d^{2h-2}) \sim \mathrm{o}(n^{2h-2})$, by the same argument, the proof is obtained. □

We consider matrix $V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}$, and we prove the following observation

Observation A.2.2. *For $0 < k \leq \frac{n}{2}$, $d(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1})$ has $2k$ number of non-zero elements, and trace$(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}) \sim O(k^2)$.*

Proof. We consider the i^{th} element of the diagonal as follows,

$$d(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1})(i) = \sum_{r=1}^{n} V_k\Sigma_n^{-1}(i,r)V_k\Sigma_n^{-1}(r,i).$$

Recall from Observation A.2.1 that there are (k+1) non-zeros elements in each column for matrix $V_k\Sigma_n^{-1}$ and, from Theorem A.2.1, $V_k\Sigma_n^{-1}(i,j) < \infty$, and thus we have

$$d(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1})(i) \sim O(k+1).$$

In addition, since for $i > 2k$, we have

$$V_k\Sigma_n^{-1}(i,j) = \begin{cases} 1 & if j = i-k \\ 0 & otherwise \end{cases},$$

and

$$V_k\Sigma_n^{-1}(i-k,i)=0,$$

we have $d(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1})(i)=0$ for $i>2k$. Thus

$$\sum_{i=1}^{n} d(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1})(i)\sim O(2k(k+1))\sim O(k^2).$$

□

We illustrate the proof as follows, for $k\leq\frac{n}{2}$

$$V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}=\begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \dots & c_{1,n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ c_{k,1} & c_{k,2} & c_{k,3} & \dots & c_{k,n}\\ 1 & 0 & 0 & \dots & 0\\ 0 & 1 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & \dots & 1 & \dots & 0 \end{bmatrix}\begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \dots & c_{1,n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ c_{d,1} & c_{d,2} & c_{d,3} & \dots & c_{k,n}\\ 1 & 0 & 0 & \dots & 0\\ 0 & 1 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & \dots & 1 & \dots & 0 \end{bmatrix}$$

$$=\left[\begin{array}{c|c} C_1 & C_2\\ \hline I_{n-k} & 0_{n\times n-k}\end{array}\right]\left[\begin{array}{c|c} C_1 & C_2\\ \hline I_{n-k} & 0_{n\times n-k}\end{array}\right]$$

with I_{n-k} the identical matrix of dimension $n-d$, and $0_{n\times n-k}$, the n by $n-k$ zero matrix.

A.2.3 Proofs on Theorem 4.5.2 and Theorem 4.5.3

Theorem 4.5.2. $\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A_k^1(n)}{n^2}\to 0$, as $n\to\infty$.

Proof. Recall from Observation A.2.1, for $k\leq n$, there are only k number of non zero elements along the diagonal and from Theorem A.2.1, the elements are finite. For $k>n$, $trace(V_k\Sigma_n^{-1})^2\sim O(n^2)\sim \mathrm{o}(k^2)$. Therefore, we have

$$\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A_k^1(n)}{n^2}=\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}trace(V_k\Sigma_n^{-1})^2}{n^2}\sim \mathrm{o}(\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}k^2}{n^2}).$$

We consider function $e^{-k\lambda_0\delta}k^2$ with $k\in[1,\infty)$. The function is increasing for

$k \in [1, \frac{2}{\lambda_0\delta})$, and is decreasing for $k \in [\frac{2}{\lambda_0\delta}, \infty)$. Let $k^* = \lfloor \frac{2}{\lambda_0\delta} \rfloor$, and μ be the counting measure,

$$\begin{aligned}
\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}k^2 &= \int_{\mathbb{N}} e^{-k\lambda_0\delta}k^2 d\mu \leq \int_{k=1}^{\infty} e^{-k\lambda_0\delta}k^2 dk + e^{-k^*\lambda_0\delta}k^{*2} \\
&= \frac{-1}{\lambda_0\delta}e^{-k\lambda_0\delta}k^2\big|_1^{\infty} + \frac{2}{\lambda_0\delta}\int_{k=1}^{\infty} e^{-k\lambda_0\delta}k dk + e^{-k^*\lambda_0\delta}k^{*2} \\
&\sim O(\frac{2}{\lambda_0^2\delta^2}e^{-k\lambda_0\delta}k^2\big|_1^{\infty} + \frac{2}{\lambda_0^2\delta^2}\int_{k=1}^{\infty} e^{-k\lambda_0\delta}dk + e^{-k^*\lambda_0\delta}k^{*2}) \\
&= \frac{2}{\lambda_0^2\delta^2}e^{-\lambda_0\delta} + \frac{2}{\lambda_0\delta^3}e^{-\lambda_0\delta} + e^{-k^*\lambda_0\delta}k^{*2} \\
&\sim O(\delta^{-3}).
\end{aligned}$$

Thus we have

$$\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A_k^1(n)}{n^2} \leq O(n^{-2}\delta^{-3}) \xrightarrow{n\to\infty} 0.$$

□

Theorem 4.5.3. $\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A^2(n)}{n^2} \to 0$ as $n \to \infty$.

Proof. Recall from Observation A.2.2, $trace(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}) \sim O(k^2)$ for $k \leq \frac{n}{2}$, and for $k > \frac{n}{2}$, $trace(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1}) \sim O(n^2) \sim \mathrm{o}(k^2)$.therefore,

$$\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}A_k^2(n)}{n^2} = \frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}trace(V_k\Sigma_n^{-1}V_k\Sigma_n^{-1})}{n^2} \sim \mathrm{o}(\frac{\sum_{k=1}^{\infty} e^{-k\lambda_0\delta}k^2}{n^2}).$$

The proof is the same as it is in Theorem 4.5.2. □

A.2.4 Calculation for (4.58)

First, we need to control $\|C_d\|_F$, and we prove a lemma first.

Lemma A.2.1. $\|\Sigma^{-\frac{1}{2}}V_{|k-l|}\Sigma^{-\frac{1}{2}}\|_F^2 \leq \frac{1}{2}(\|V_{|k-l|}\Sigma_n^{-1}\|_F^2 + \|\Sigma_n^{-1}V_{|k-l|}\|_F^2)$

Proof. Let A be an n by n matrix and we have [44]

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{j=0}^{n} A(i,j)^2 = trace(A^*A),$$

and therefore we have $\|UA\|_F^2 = \|UA\|_F^2 = \|A\|_F^2$, where $U^*U = UU^* = I$ with I the identical matrix [44] .

We consider the eigenvalue decomposition of $\Sigma_n^{-1} = U^*\Lambda U$, where U is orthonormal and Λ is the eigenvalue matrix. Thus, $\Sigma_n^{-\frac{1}{2}} = U^*\Lambda^{\frac{1}{2}}U$, and we have the following

$$\begin{aligned}
\|C_d\|_F^2 &= \|\Sigma_n^{-\frac{1}{2}} V_d \Sigma_n^{-\frac{1}{2}}\|_F^2 = \|U^*\Lambda^{\frac{1}{2}}UV_dU^*\Lambda^{\frac{1}{2}}U\|_F^2 \\
&= \|\Lambda^{\frac{1}{2}}UV_dU^*\Lambda^{\frac{1}{2}}U\|_F^2 = \|\Lambda^{\frac{1}{2}}UV_dU^*\Lambda^{\frac{1}{2}}\|_F^2 \\
&\leq \frac{1}{2}(\|UV_dU^*\Lambda\|_F^2 + \|\Lambda UV_dU^*\|_F^2) \qquad \text{(a)} \\
&= \frac{1}{2}(\|V_dU^*\Lambda U\|_F^2 + \|U^*\Lambda UV_d\|_F^2) \\
&= \frac{1}{2}(\|V_d\Sigma_n^{-1}\|_F^2 + \|\Sigma_n^{-1}V_d\|_F^2),
\end{aligned}$$

where step (a) is due to the following, for any matrix B, we have

$$|\Lambda B\Lambda(i,j)| = |B(i,j)\lambda_i\lambda_j| \leq \frac{1}{2}|B(i,j)|(\lambda_i^2 + \lambda_j^2) = \frac{1}{2}(\Lambda^2 B + B\Lambda^2)(i,j).$$

□

Then, we prove that

$$S_{1,2} \sim O(\frac{1}{n(1-e^{-\lambda_0\delta})^3}).$$

Proof. Since by Observation A.2.1, for $i > d$ and $d < n$, we have

$$V_d\Sigma_n^{-1}(i,j) = \begin{cases} 1, & j = i-d \\ 0, & otherwise \end{cases},$$

and similarly, we have for $1 \leq j \leq (n-d)$,

$$\Sigma_n^{-1}V_d(i,j) = \begin{cases} 1, & i = d+j \\ 0, & otherwise \end{cases}.$$

That is for matrices as $V_d\Sigma_n^{-1}$ ($\Sigma_n^{-1}V_d$), the i^{th} row (column) has only one non-zero element for $i > d$. From Theorem A.2.1, $\Sigma_n^{-1}V_d(i,j) < \infty$, and thus we have

$$\|C_d\|_F^2 = \begin{cases} (n-d) + O(nd) & d \leq n-1 \\ O(n^2) & d \geq n \end{cases}.$$

Since $n^2 \leq dn$ for $d \geq n$,

$$\begin{aligned} \frac{2}{n^2}\left(\sum_{k=1}^{\infty}\sum_{d=1}^{\infty} e^{-(2k+d)\lambda_0\delta}\|C_d\|_F^2\right) &\sim \mathrm{o}(\frac{2}{n^2}\sum_{k=0}^{\infty}\sum_{d=1}^{\infty} e^{-(2k+d)\lambda_0\delta}(n-d) + d \times n) \\ &\sim (\frac{2}{n}\sum_{k=0}^{\infty}\sum_{d=1}^{\infty} e^{-(2k+d)\lambda_0\delta}d), \end{aligned}$$

and by the summation by parts [58], we have

$$\begin{aligned} &\frac{2}{n}\frac{1}{1-e^{-2\lambda_0\delta}}\sum_{d=1}^{\infty} e^{-d\lambda_0\delta}d \\ &= \frac{2}{n}\frac{1}{1-e^{-2\lambda_0\delta}}\lim_{m\to\infty}\sum_{d=1}^{m}\frac{-e^{-\lambda_0\delta}(1-e^{-d\lambda_0\delta})}{1-e^{-\lambda_0\delta}} + m\frac{e^{-\lambda_0\delta}(1-e^{-m\lambda_0\delta})}{1-e^{-\lambda_0\delta}} \\ &= \frac{2}{n}\frac{1}{1-e^{-2\lambda_0\delta}}\frac{e^{-\lambda_0\delta}}{1-e^{-\lambda_0\delta}}\lim_{m\to\infty}\sum_{d=1}^{m} e^{-d\lambda_0\delta} \\ &\sim O(\frac{1}{n(1-e^{-\lambda_0\delta})^3}). \end{aligned}$$

Therefore, the result is proved. □

A.3 Calculations for section 4.5.6

A.3.1 Proof for (4.73)

We would like to prove the equation (4.73),

$$\mathbb{E}\left(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}\int_0^{\delta} u dX_{t_{n-j}-u}\right)^2 = \mathbb{E}\left(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}\int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du\right)^2.$$

We first prove a lemma

Lemma A.3.1. $\mathbb{E}(\int_0^\delta u dX_{t_i-u} - \int_0^\delta X_{t_{i-1}} - X_{t_i-u} du)^2 = 0.$

Proof. Since the increments of a fractional Brownian motion are stationary, we prove the lemma at $t_i = 0$, that is

$$\mathbb{E}(\int_0^\delta u dX_{-u} - \int_0^\delta X_{-\delta} - X_{-u} du)^2 = 0.$$

Recall from Theorem 8, the fractional integral $\int_0^\delta u dX_{-u}$ is defined as follows,

$$\int_0^\delta u dX_{-u} = C(h) \int_{\mathbb{R}} I_-^{h-\frac{1}{2}}(u 1_{[0,\delta]})(t) dW_{-t},$$

where $C(h) = (2h sin\pi h \Gamma(2h))^{\frac{1}{2}}$, $(I_-^\alpha f)(x) = \frac{1}{\Gamma(\alpha)} \int_x^\infty f(s)(s-x)^{\alpha-1} ds$, and $\{W_t\}$ is a Brownian motion on $(\Omega, \mathcal{F}, P)$.
For $h > 0.5$, by Theorem 4.1.2, $\{C(h) I_-^{h-\frac{1}{2}}(\mathbf{1}_{[u,v)}), u, v \in \mathbb{R}\}$ is dense in $L_2(\mathbb{R})$, and by Theorem 4.1.1, we have $C(h) I_-^{h-\frac{1}{2}}(u 1_{[0,\delta]})(.) \in L_2(\mathbb{R})$, and thus

$$\lim_{m\to\infty} \int_{\mathbb{R}} (C(h) I_-^{h-\frac{1}{2}}(u_m 1_{[0,\delta]} - u 1_{[0,\delta]}))^2(t) dt = 0$$

where u_m is a step function as follows

$$u_m(x) = \sum_{i=0}^{m-1} \frac{i}{m} \delta \mathbf{1}_{[\frac{i}{m}\delta, \frac{i+1}{m}\delta)}(x).$$

Therefore,

$$\lim_{m\to\infty}\int_{\mathbb{R}}(C(h)I_{-}^{h-\frac{1}{2}}(u_m 1_{[0,\delta]}-u1_{[0,\delta]}))^2(t)dt$$
$$=\lim_{m\to\infty}\mathbb{E}(C(h)\int_{\mathbb{R}}I_{-}^{h-\frac{1}{2}}(u1_{[0,\delta]}-u_m 1_{[0,\delta]})(t)dW_{-t})^2$$
$$=\lim_{m\to\infty}\mathbb{E}(\int_0^\delta t-u_m(t)dX_{-t})^2=\lim_{m\to\infty}\mathbb{E}\left(\int_0^\delta tdX_{-t}-(X_{-\delta}u_m(\delta)-\int_0^\delta X_{-t}du_m(t))\right)^2. \tag{A.4}$$

Since $X_{-\delta}u_m(\delta)-\int_0^\delta X_{-t}du_m(t)=\int_0^\delta u_m(t)dX_{-t}$, and $\mathbb{E}(\int_0^\delta u_m(t)dX_{-t})^2=\int_0^\delta\int_0^\delta u_m(t)u_m(s)|s-t|^{2h-2}dtds<\delta^{2h}$ [46] (Theorem 1.7.3, p25), by Dominated Convergence Theorem (A.1.5), we can interchange the limit operation and the integral in (A.4) as follows,

$$\lim_{m\to\infty}\int_{\mathbb{R}}(C(h)I_{-}^{h-\frac{1}{2}}(u_m 1_{[0,\delta]}-u1_{[0,\delta]}))^2(t)dt$$
$$=\lim_{m\to\infty}\mathbb{E}(\int_0^\delta tdX_{-t}-(X_{-\delta}u_m(\delta)-\int_0^\delta X_{-t}du_m(t)))^2$$
$$=\mathbb{E}(\int_0^\delta tdX_{-t}-(X_{-\delta}\delta-\int_0^\delta X_{-t}dt))^2.$$

For $h<0.5$, we use the result in [46] (Chapter 1.8, p29) where Mishura and Yuliya proved that for $f:[0,T]\to\mathbb{R}$, and $\{B_t^h\}$ an h-fractional Brownian motion on $(\Omega,\mathcal{F},P)$, if f is of bounded variation and is continuous, then we have

$$\mathbb{E}(\int_0^T f(t)dB_t^h-f(T)B_T^h-\int_0^T B_t^h df(t))^2=0.$$

Since, $f(x)=x$ is continuous and of bounded variation on $[0,\delta]$, then by the above result, we have the proof for $h<\frac{1}{2}$. □

Next, we prove

$$\mathbb{E}\left(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}\int_0^{\delta} u dX_{t_{n-j}-u}\right)^2 = \mathbb{E}\left(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}\int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du\right)^2.$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left(\sum_{j=1}^{\infty} e^{(-j+1)\delta\lambda_0}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)\right)^2 \\
&= \mathbb{E}\sum_{j=0}^{\infty} e^{-2j\delta\lambda_0}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)^2 \\
&+ 2\mathbb{E}\sum_{j=0}^{\infty}\sum_{k>j}^{\infty} e^{-(j+k)\delta\lambda_0}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)(\int_0^{\delta} u dX_{t_k+u} - \int_0^{\delta} X_{t_{k+1}} - X_{t_k+u}du) \\
&\leq \mathbb{E}\sum_{j=0}^{\infty} e^{-2j\delta\lambda_0}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)^2 \\
&+ 2\mathbb{E}\sum_{j=0}^{\infty}\sum_{k>j}^{\infty} e^{-(j+k)\delta\lambda_0}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)^2 \\
&= \sum_{j=0}^{\infty} e^{-2j\delta\lambda_0}\mathbb{E}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)^2 \\
&+ 2\sum_{j=0}^{\infty}\sum_{k>j}^{\infty} e^{-(j+k)\delta\lambda_0}\mathbb{E}(\int_0^{\delta} u dX_{t_{n-j}-u} - \int_0^{\delta} X_{t_{n-j-1}} - X_{t_{n-j}-u}du)^2,
\end{aligned}$$

where the second last step is due to the Monotonic Convergence Theorem (A.1.11). By Lemma A.3.1, the above equation is zero. □

A.3.2 The order for $Q_{2,2}(\delta)$

We prove the observation as follows,

Observation A.3.1. $\mathbb{E}(Q_{2,2}(\delta)) \sim O(\delta^3 \sum_{d=1}^{\infty} e^{-d\lambda_0\delta}(d\delta)^{2h-2})$.

Proof. We consider $Q_{2,2}(\delta)$, and since the increments of the fractional Brownian

motion are stationary, we consider the following, for $d = k - j$,

$$\mathbb{E}\left(\int_0^\delta\int_0^\delta (X_{t_0} - X_{t_1-u})(X_{t_{-d}} - X_{t_{-d+1}-v})dudv\right).$$

Since

$$\int_0^\delta\int_0^\delta \mathbb{E}|(X_{t_0} - X_{t_1-u})(X_{t_{-d}} - X_{t_{-d+1}-v})|dudv \le \int_0^\delta\int_0^\delta \mathbb{E}(X_{t_0} - X_{t_1-u})^2 dudv = c\delta^{2h+2},$$

by Fubini's theorem (A.1.3), we have

$$\begin{aligned}
&\mathbb{E}\left(\int_0^\delta\int_0^\delta (X_{t_0} - X_{t_1-u})(X_{t_{-d}} - X_{t_{-d+1}-v})dudv\right)\\
&= \int_0^\delta\int_0^\delta \mathbb{E}\left((X_{t_0} - X_{t_1-u})(X_{t_{-d}} - X_{t_{-d+1}-v})\right)dudv\\
&= \frac{1}{2}\int_0^\delta\int_0^\delta \left(|d\delta + (v-\delta)|^{2h} + |d\delta - (u-\delta)|^{2h} - |d\delta + (v-u)|^{2h} - |d\delta|^{2h}\right)dudv\\
&= \frac{1}{2}\int_0^\delta\int_0^\delta \left((d\delta + (v-\delta))^{2h} + (d\delta - (u-\delta))^{2h} - (d\delta + (v-u))^{2h} - d\delta^{2h}\right)dudv.
\end{aligned}$$

By applying Taylor expansion (A.1.1) to the integrands above at $d\delta + \frac{v-u}{2}$, we have

$$\begin{aligned}
&(d\delta + v - \delta)^{2h} = (d\delta + \frac{v-u}{2} + \frac{v+u}{2} - \delta)^{2h}\\
&\sim O\left((d\delta + \frac{v-u}{2})^{2h} + 2h(d\delta + \frac{v-u}{2})^{2h-1}(\frac{v+u}{2} - \delta) + h(2h-1)(d\delta + \frac{v-u}{2})^{2h-2}(\frac{(v+u)}{2} - \delta)^2\right)
\end{aligned}$$

$$\begin{aligned}
&(d\delta - u + \delta)^{2h} = (d\delta + \frac{v-u}{2} - (\frac{v+u}{2} - \delta))^{2h}\\
&\sim O\left((d\delta + \frac{v-u}{2})^{2h} - 2h(d\delta + \frac{v-u}{2})^{2h-1}(\frac{v+u}{2} - \delta) + h(2h-1)(d\delta + \frac{v-u}{2})^{2h-2}(\frac{(v+u)}{2} - \delta)^2\right)
\end{aligned}$$

$$(d\delta + v - u)^{2h} = (d\delta + \frac{v-u}{2} + \frac{v-u}{2})^{2h}$$

$$\sim O\left((d\delta+\frac{v-u}{2})^{2h}+2h(d\delta+\frac{v-u}{2})^{2h-1}(\frac{v-u}{2})+h(2h-1)(d\delta+\frac{v-u}{2})^{2h-2}(\frac{(v-u)}{2})^2\right),$$

and

$$(d\delta)^{2h}=(d\delta+\frac{v-u}{2}-\frac{v-u}{2})^{2h}$$
$$\sim O\left((d\delta+\frac{v-u}{2})^{2h}-2h(d\delta+\frac{v-u}{2})^{2h-1}(\frac{v-u}{2})+h(2h-1)(d\delta+\frac{v-u}{2})^{2h-2}(\frac{(v-u)}{2})^2\right).$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}\left(\int_0^\delta\int_0^\delta(X_{t_0}-X_{t_1-u})(X_{t_{-d}}-X_{t_{-d+1}-v})dudv\right)\\
&\sim\frac{1}{2}O\left(\int_0^\delta\int_0^\delta(d\delta+\frac{v-u}{2})^{2h-2}(\frac{(u+v-\delta)^2}{2}-\frac{(u-v)^2}{2})dudv\right)\\
&\sim(d\delta-\frac{\delta}{2})^{2h-2}\frac{1}{2}\mathrm{o}\left(\int_0^\delta\int_0^\delta 2uvdudv-2\int_0^\delta\int_0^\delta\delta(u+v)dudv+\int_0^\delta\int_0^\delta\delta^2dudv\right)\\
&\sim O((d\delta-\frac{\delta}{2})^{2h-2}\delta^4).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{E}(Q_{2,2}(\delta))\\
&=\mathbb{E}\left(\sum_{j=0}^\infty\sum_{d=1}^\infty e^{-2j\lambda_0\delta}e^{-d\lambda_0\delta}\int_0^\delta(X_{t_{n-i-j-1}}-X_{t_{n-i-j-u}})du\int_0^\delta(X_{t_{-i-j-d+1}}-X_{t_{-i-j-d+u}})du\right)\\
&=\sum_{j=0}^\infty\sum_{d=1}^\infty e^{-2j\lambda_0\delta}e^{-d\lambda_0\delta}O((d\delta-\frac{\delta}{2})^{2h-2}\delta^4)\sim O(\delta^3\sum_{d=1}^\infty e^{-d\lambda_0\delta}(d\delta)^{2h-2}).
\end{aligned}$$

□

Bibliography

[1] Aït-Sahalia and Yacine. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.

[2] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 03 2001.

[3] Krishna B. Athreya and Soumendra N. Lahiri. *Measure Theory and Probability Theory*. Springer, New York, 2006.

[4] Fristedt B. and Gray L. *A Modern Approach to Probability Theory*. Birkhäuser Basel, Boston, MA, 1997.

[5] Ole Barndorff-Nielsen and Neil Shephard. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477, Sep 2002.

[6] Heinz H Bauschke, JY Cruz, Tran TA Nghia, Hung M Phan, and Xianfu Wang. Optimal rates of convergence of matrices with applications. *arXiv preprint arXiv:1407.0671*, 2014.

[7] Rachid Belfadli, Khalifa Es-Sebaiy, and Youssef Ouknine. Parameter estimation for fractional Ornstein-Uhlenbeck processes: non-ergodic case. *arXiv preprint arXiv:1102.5491*, 2011.

[8] F Jay Breidt, Nuno Crato, and Pedro De Lima. The detection and estimation of long memory in stochastic volatility. *Journal of econometrics*, 83(1-2):325–348, 1998.

[9] Alexandre Brouste and Stefano M Iacus. Parameter estimation for the discretely observed fractional Ornstein–Uhlenbeck process and the Yuima R package. *Computational Statistics*, 28(4):1529–1547, 2013.

[10] F Bünger. Inverses, determinants, eigenvalues, and eigenvectors of real symmetric Toeplitz matrices with linearly increasing entries. *Linear Algebra and its Applications*, 459:595–619, 2014.

[11] V. Butkus and R. Norvaiša. Computation of p-variation. *Lithuanian Mathematical Journal*, 58(4):360–378, 2018.

[12] Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab*, 8(3):1–14, 2003.

[13] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. *arXiv preprint arXiv:1603.03788*, 2016.

[14] Pavel Chigansky and Marina Kleptsyna. Exact asymptotics in eigenproblems for fractional Brownian covariance operators. *Stochastic Processes and their Applications*, 128(6):2007 – 2059, 2018.

[15] Alexandra Chronopoulou and Frederi G Viens. Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quantitative Finance*, 12(4):635–649, 2012.

[16] Thomas H Cormen. *Introduction to algorithms*. MIT press, Cambridge, Massachusetts, 2009.

[17] Laure Coutin and Zhongmin Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probability theory and related fields*, 122(1):108–140, 2002.

[18] Hui Dai, Zachary Geary, and Leo P Kadanoff. Asymptotics of eigenvalues and eigenvectors of Toeplitz matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(05):P05012, 2009.

[19] Andrew S Dickinson. Optimal approximation of the second iterated integral of Brownian motion. *Stochastic Analysis and Applications*, 25(5):1109–1128, 2007.

[20] Joscha Diehl, Peter Friz, and Hilmar Mai. Pathwise stability of likelihood estimators for diffusions via rough paths. *arXiv preprint arXiv:1311.1061*, 2013.

[21] Reinhard Diestel. *Graph theory*. Springer Publishing Company, Incorporated, 2018.

[22] Bruce K Driver. Rough path analysis. *Online notes available at http://www. math. ucsd. edu/bdriver*.

[23] Richard M Dudley and Rimas Norvaiša. *Differentiability of six operators on nonsmooth functions and p-variation*. Springer, Berlin, 2006.

[24] Khalifa Es-Sebaiy. Berry–esséen bounds for the least squares estimator for discretely observed fractional ornstein–uhlenbeck processes. *Statistics & Probability Letters*, 83(10):2372–2385, 2013.

[25] Ronald Gallant et al. Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, 84(1):125–141, 1997.

[26] Stephen Figlewski, William L Silber, and Marti G Subrahmanyam. *Financial options: from theory to practice*. Irwin Professional Pub, New York, 1990.

[27] Peter K Friz and Nicolas B Victoir. *Multidimensional stochastic processes as rough paths: theory and applications*, volume 120. Cambridge University Press, Cambridge, 2010.

[28] Benjamin Graham. Sparse arrays of signatures for online character recognition. *arXiv preprint arXiv:1308.0371*, 2013.

[29] Ulf Grenander and Gabor Szegö. *Toeplitz forms and their applications*. AMS Chelsea Publishing, New York, 1984.

[30] Maria Letizia Guerra and Luciano Stefanini. A comparative simulation study for estimating diffusion coefficient. *Mathematics and Computers in Simulation*, 53(3):193 – 203, 2000.

[31] Alemdar Hasanov Hasanoğlu and Vladimir G Romanov. *Introduction to Inverse Problems for Differential Equations*. Springer, Cham, Switzerland, 2017.

[32] Yaozhong Hu and David Nualart. Parameter estimation for fractional ornstein–uhlenbeck processes. *Statistics & Probability Letters*, 80(11):1030–1038, 2010.

[33] Yaozhong Hu and Jian Song. Parameter estimation for fractional ornstein–uhlenbeck processes with discrete observations. In *Malliavin Calculus and Stochastic Analysis*, pages 427–442. Springer, Boston, MA, 2013.

[34] Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, New York, 2013.

[35] ML Kleptsyna, A Le Breton, and M-C Roubaud. Parameter estimation and optimal filtering for fractional type stochastic systems. *Statistical inference for stochastic processes*, 3(1-2):173–182, 2000.

[36] Dirk P Kroese and Zdravko I Botev. Spatial process generation. *arXiv preprint arXiv:1308.0399*, 2013.

[37] Alexander Kukush, Yulia Mishura, and Esko Valkeila. Statistical inference with fractional Brownian motion. *Statistical inference for stochastic processes*, 8(1):71–93, 2005.

[38] José R León and Carenne Ludeña. Estimating the diffusion coefficient for diffusions driven by fbm. *Statistical inference for stochastic processes*, 3(1-2):183–192, 2000.

[39] Daniel Levin, Terry Lyons, and Hao Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *arXiv:1309.0260*, 2016.

[40] Terry Lyons, Hao Ni, and Harald Oberhauser. A feature set for streams and an application to high-frequency financial tick data. In *Proceedings of the 2014 International Conference on Big Data Science and Computing*, page 5, New York, 2014. ACM.

[41] Terry J Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Springer, Heidelberg, 2007.

[42] Bo Martin Bibby and Michael Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2):17–39, 03 1995.

[43] Hiroyuki Matsumoto and Setsuo Taniguchi. *Stochastic analysis: Itô and Malliavin calculus in tandem*, volume 159. Cambridge University Press, Cambridge, 2016.

[44] Carl D Meyer. *Matrix analysis and applied linear algebra*. Siam, Philadelphia, 2000.

[45] Nichols J.M. Bucholtz F. et al. Michalowicz, J.V. An Isserlis' Theorem for mixed Gaussian variables: Application to the auto-bispectral density. *Journal of Statistical Physics*, 36(1):89–102, 2009.

[46] Yuliya Mishura. *Stochastic calculus for fractional Brownian motion and related processes*, volume 1929 of *Lecture Notes in Mathematics*. Springer Science & Business Media, Heidelberg, 2008.

[47] Yuliya Mishura, Vladimir I Piterbarg, Kostiantyn Ralchenko, and Anton Yurchenko-Tytarenko. Stochastic representation and pathwise properties of fractional cox-ingersoll-ross process. *arXiv preprint arXiv:1708.02712*, 2017.

[48] PA Mykland and L Zhang. The econometrics of high frequency data. In *Statistical Methods for Stochastic Differential Equations*, pages 109–190. Chapman & Hall/CRC P ress, New York, 2012.

[49] David Nualart. *The Malliavin calculus and related topics*, volume 1995 of *Probability, its Applications*. Springer, Berlin, Heidelberg, 2006.

[50] Anastasia Papavasiliou. Coarse-grained modeling of multiscale diffusions: the p- variation estimates. *Stochastic Analysis 2010*, 02 2010.

[51] Anastasia Papavasiliou and Christophe Ladroue. Parameter estimation for rough differential equations. *The Annals of Statistics*, 39(4):2047–2073, 2011.

[52] Anastasia Papavasiliou and Kasia B Taylor. Approximate likelihood construction for rough differential equations. *arXiv preprint arXiv:1612.02536*, 2016.

[53] Stuart Pavliotis, G.A. Parameter estimation for multiscale diffusions. *Journal of Statistical Physics*, 127(741-748), 2007.

[54] Asger Roer Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian journal of statistics*, 22:55–71, 1995.

[55] BLS Prakasa Rao. Conditions for singularity for measures generated by two fractional psuedo-diffusion processes. *Stochastic Analysis and Applications*, 34(2):183–192, 2016.

[56] Zhongmin Qian and Xingchen Xu. Lévy area of fractional Ornstein-Uhlenbeck process and parameter estimation. *arxiv:1803.11039v2*, 2018.

[57] Gareth O Roberts and Osnat Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika*, 88(3):603–621, 2001.

[58] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill.Inc, New York, 1976.

[59] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24 of *Algorithms and Combinatorics*. Springer, Berlin, Heidelberg, 2003.

[60] Stefano Serra. On the extreme eigenvalues of Hermitian (block) Toeplitz matrices. *Linear algebra and its applications*, 270(1-3):109–129, 1998.

[61] Helle Sørensen. Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, 72(3):337–354, 2004.

[62] V. S. Sunder. *Operators on Hilbert space*, volume 71 of *Texts and Readings in Mathematics*. Springer, Singapore, 2016.

[63] Anastasia Papavasiliou Theodoros Manikas. Diffusion parameter estimation for the homogenized equation. *arxiv:1907.00915*, 2018.

[64] Ciprian A Tudor and Frederi G Viens. Statistical aspects of the fractional stochastic calculus. *The Annals of Statistics*, 35:1183–1212, 2007.

[65] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

[66] Darren J Wilkinson. *Stochastic modelling for systems biology*, volume 8. Chapman and Hall/CRC, London, 2006.

[67] David Williams. *Probability with martingales*. Cambridge University Press, Cambridge, 1991.

[68] Antoni Zygmund. *Trigonometric series*, volume 1. Cambridge university press, Cambridge, 2002.